SPSS: Stats Practically Short and Simple

Sidney Tyrrell



Download free books at



Sidney Tyrrell

SPSS: Stats Practically Short and Simple

SPSS: Stats Practically Short and Simple © 2009 Sidney Tyrrell & Ventus Publishing ApS ISBN 978-87-7681-474-8

Contents

| 1. An Overview | 8 |
|---------------------------------|----|
| Getting In | 8 |
| Frequencies | 9 |
| Exporting your Output to Word | 13 |
| Drawing charts | 14 |
| Exercise | 14 |
| Moving Around | 15 |
| 2. Entering Data | 16 |
| Introduction | 16 |
| Entering Data directly | 16 |
| Defining Variables | 17 |
| Adjusting the width | 17 |
| Variable names | 18 |
| Entering data via a spreadsheet | 19 |
| Adding Variable Labels | 19 |
| Adding Value Labels | 20 |
| Important note | 21 |
| Finally | 21 |
| 3. Editing and Handling Data | 22 |
| Correcting entries | 22 |

And and

The next step for top-performing graduates

Please click the advert



Contents

| Deleting entries | 22 |
|---|----|
| Copying cells, columns and rows | 22 |
| Inserting a variable (a column) | 22 |
| Inserting a case (a row) | 23 |
| Moving columns | 23 |
| Sorting data | 23 |
| Saving data and output | 23 |
| Exporting Output | 23 |
| Saving Data as an Excel file | 24 |
| Copying tables and charts into Word | 24 |
| Printing from SPSS | 24 |
| Recoding into groups | 25 |
| Revision exercise | 26 |
| Doing Calculations on Variables | 27 |
| Selecting a subset | 28 |
| Selecting a Random Sample | 29 |
| Merging Files | 31 |
| Adding Variables | 31 |
| Adding cases | 32 |
| 1 Description Statistics | 22 |
| 4. Descriptive Statistics | 33 |
| The Functions | 34 |
| Finding Frequencies for Multiple Response Variables | 37 |
| Tables are tricky! | 44 |



u.s. 1-800-829-4444 canada: 1-877-894-4414

© Agilent Technologies, Inc. 2012

Please click the advert

Anticipate ____Accelerate ____Achieve



Agilent Technologies

© UBS 2010. All rights reserved.

| 5. Charts | 46 |
|---|----|
| Introduction | 46 |
| A Simple Bar Chart | 47 |
| A clustered bar chart | 50 |
| Percentage Clustered Bar Chart using Legacy Dialogs | 51 |
| With correct labels! | 51 |
| A stacked % bar chart | 53 |
| Drawing a panel bar chart | 53 |
| Drawing a bar chart of more than one variable | 54 |
| Drawing a pie chart | 56 |
| Histogram | 58 |
| Boxplots | 60 |
| 6. Regression and Correlation | 63 |
| Introduction | 63 |
| Scatter Diagrams | 64 |
| Correlation | 64 |
| Correlation and Causation | 65 |
| Regression | 65 |
| Multiple Regression | 68 |
| 7. Statistical Tests | 70 |
| The One-Sample T test | 70 |
| The Chi-Squared Test for contingency tables | 72 |
| t-test for related samples | 72 |
| | |



You're full of *energy and ideas*. And that's just what we are looking for.

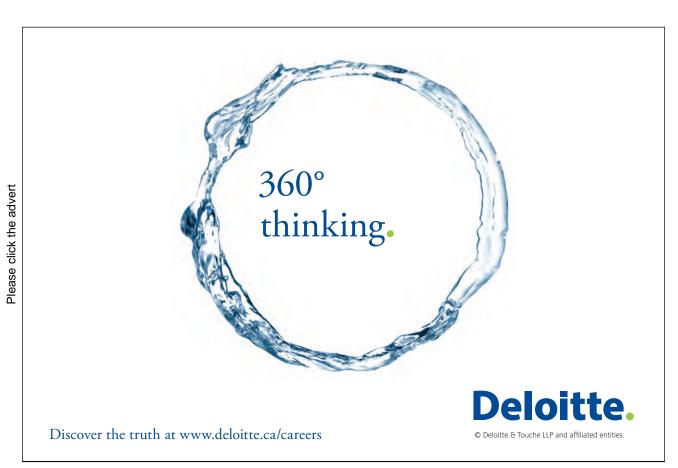
Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting www.ubs.com/graduates.

www.ubs.com/graduates



| 8. And finally | 83 |
|--|----|
| Wilcoxon Signed-Ranks test for paired samples | 81 |
| Non-Parametric Tests | 79 |
| Analysis of Variance | 78 |
| t-test for the differences in the Means of independent samples | 77 |



1. An Overview

Getting In

Having opened SPSS you will get a dialogue box which you can cancel the first time you enter SPSS. Enlarge the window.

SPSS is like a spreadsheet **but it does** not update calculations, tables or charts if you change the data.

At the top of the screen are a series of menus which can be used to instruct SPSS to do something.

| 🖬 U | 🛂 Untitled1 [DataSet0] - SPSS Data Editor | | | | | | | | | | | |
|--------------|---|--------------|--------------|-------------------|-----------------|----------------|-------------------|------------------|--------|--------------|--|--|
| <u>F</u> ile | <u>E</u> dit | <u>∨</u> iew | <u>D</u> ata | <u>T</u> ransform | <u>A</u> nalyze | <u>G</u> raphs | <u>U</u> tilities | Add- <u>o</u> ns | Window | <u>H</u> elp | | |
| ₿ I | | ШŤ | •• | | ? M | + | # | iii 🥳 🤅 | è 🗣 | | | |

SPSS uses 2 windows: The Data Editor, which is what you are looking at and which has 2 tabs at the bottom, and the Viewer.

The Viewer is not visible yet, but opens automatically as soon as you open a file or run a command that produces output, such as statistics, tables and charts.

The menus are the same in each window but the icons are different. To switch between the two windows use the tabs at the bottom of the screen.

The Data Editor window:

| | 🖼 Untitled1 [DataSetU] - SPSS Data Editor | | | | | | | | | | | | | |
|----------------|---|-------|------------------------------|--------------------|----------------------------|-------------|-------|------------|--------------------|----------------------|----|------------------|---|----------------------------|
| File | Ē | dit | ⊻iew | <u>D</u> ata | <u>T</u> ransform | <u>A</u> na | alyze | <u>G</u> r | aphs | Utilities | Ac | d- <u>o</u> ns | У | <u>M</u> indo [,] |
| ß | Ŗ | | ; | •• | - i | ? | 44 | + | | - | | \$ | Ø | |
| Open S File | Save | Print | Review recent dialogue | Undo Redo boxes | Go to Go to case variat | | Find | | Insert Variable | Split Weigh Cases | | Show labels S | | Show All |

The Output window:

| | *Οι | ıtp | ut1 | [Docu | ument 1 | [] - SPS | SS Viev | ver | | | |
|--------------|---------------|----------------|------|--------------------------------------|---|--|------------------------------------|-----------------|----------------|-----------|-------|
| Eile | Ed | lit | View | <u>D</u> ata | Transfor | m <u>I</u> nsert | Format | <u>A</u> nalyze | <u>G</u> raphs | Utilities | Add-g |
| | | D. | Q. | 🕒 🗉 | | Image: A test and a second | | ି କ | <u>, a</u> | i 🖌 🗗 | - @P |
| Open File | Save Print | print Print | re | xport Recall l cent ogue boxes | Jndo Redo Go case vari Go to data | to Go to Variables able Case | Select LastShow Output Variable | | labels Sets | | |

SPSS comes with a large number of sample data files, which this book will use. If you do not have access to these, use any data set you have access to.

To open the data file **1991 U.S. General Social Survey.sav** use File > Open > Data

- Double click on the appropriate directories to open each
- Double click on the file 1991 U.S. General Social Survey.sav

At first you will probably be faced by a mass of seemingly meaningless numbers.

If you look along the toolbar you will find the Value labels icon S. Click on this and the output should look more friendly.

| | | | | | | | | | |
|--------------|---------|----|-------------------|-----------------|----------------|-----------|---------|----------------|--------------|
| <u>F</u> ile | | | <u>T</u> ransform | <u>A</u> nalyze | <u>G</u> raphs | Utilities | Add-ons | <u>Wi</u> ndow | <u>H</u> elp |
| B | | •• | 1 III- | ? M | + | 🔡 🦺 | 📰 😵 🤇 | ð 🌑 | |

• Click on the Variables icon **[?** to get an overview of each variable.

| Variable Information: | | | | | | | | | | | |
|---|--------|---|--|--|--|--|--|--|--|--|--|
| V Variable Image: Respondent's Sex [sex] Resce of Respondent [ra Image: Region of the United Sta Region of the United Sta Image: Region of the United Sta Ceneral Happiness [hap Image: Region of the Exciting or Dull [life] Number of Brothers and Image: Number of Children [chil Age of Respondent [age] Image: Highest Year School Co Highest Year School Co | ▲ ※ | sex Label: Respondent's Sex Type: F1 Missing Values: none Measurement Level: Nominal Value Labels: 1 Male 2 Female | | | | | | | | | |

Exercise:

• How many Regions of the United States are represented?

Frequencies

- Let's start simply. All that data looks a bit overwhelming so we need to get a handle on it and pick out the main messages.
- First of all how many men and women are there in this group?

For a simple count, and for percentages use

Analyze > Descriptive Statistics > Frequencies .

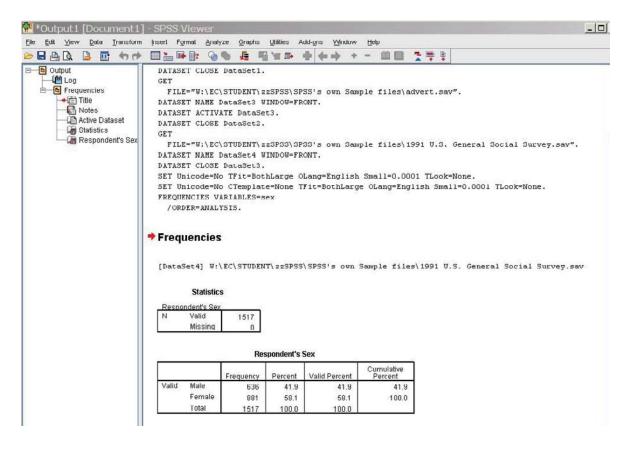
SPSS uses Dialogue boxes for the selection of variables and options.

The source list contains the list of variables, with icons as before indicating data types.

- Your dialogue box may have only listed the variable **names**, e.g. *sex*, rather than the variable **labels** such as 'Respondent's sex'. It is more helpful in analysis to see these labels.
- If they are not shown use **Edit** > **Options**
- Select the General tab and at the top under Variable Lists click on the circle Display Labels.

| Use the arrow button to move a variable to the target list – the Variable(s) box on the right. | Frequencies Variable(s): Statistics Charts |
|--|--|
| Place Respondent's sex in the Variable(s) box | Region of the United Constal Happiness [Is Life Excling or Dull Number of Brothers Number of Children [Age of Respondent [Highest Year of Sch |
| then click on OK | Display frequency tables OK Paste Reset Cancel Help |

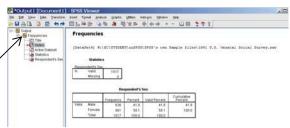
The resulting output introduces us to the Viewer window, and shows that 636 respondents, or 42%, were men. **Maximise the Viewer window.**



- There is a lot of clutter here.
- Tip: Always delete unnecessary Output, and annotate the rest as you go.
- Click on all the text at the top of the screen and press Delete on your keyboard.

The left hand pane contains the outline view. To go directly to an item click on it; very useful when you have masses of output. If you don't need it all for the moment you can hide it by clicking on the minus signs that appear in the left hand frame.

To hide one item, click on the minus sign. This is useful if you only want to print a small selection, as only what is shown is printed.



To change the order in which the items are displayed, drag and drop in the left hand pane. Try it.

To delete an item, click on it and press delete.

- Tip: Never do any analysis without interpreting it.
- To annotate your output use **Insert > New Text which** provides a text box in which you can write a comment.
- It appears on the left hand side of the screen with a red arrow at first
- Click on it and the box will open in the right hand pane for you to write in.

| Dutput | FREQUENCIES VARIABLES=sex /order=analysis. |
|--|---|
| Title TopSS Text Title TopSS Text Title Title TopSS Text Title T | Frequencies |
| └──(â) Respondent's Sex | [DataSet4] W:\EC\STUDENT\zzSP |
| | Statistics |
| | Respondent's Sex N Valid 1517 Missing 0 |

- Back to the output: itself; this can be edited.
- **Double click** on the table to bring up the Formatting Toolbar.
- If it does not appear use View>Toolbar

| M SP | SS Pivot | : Table Re | sponder | nt's Sex | _ | | | | | | |
|--|---------------------------|----------------------|--------------------------|--------------|-------------------------------------|--|--|--|--|--|--|
| Eile | <u>E</u> dit <u>V</u> iew | Insert <u>P</u> ivot | F <u>o</u> rmat <u>I</u> | <u>H</u> elp | | | | | | | |
| | | Res | spondent's | Sex | | | | | | | |
| Frequency Percent Valid Percent Cumulative | | | | | | | | | | | |
| Valid | Male | 636 | 41.9 | 41.9 | 41.9 | | | | | | |
| | Female | 881 | 58.1 | 58.1 | 100.0 | | | | | | |
| | Total | 1517 | 100.0 | 100.0 | | | | | | | |
| | | | | | | | | | | | |
| Formatting Toolbar | | | | | | | | | | | |
| • | 🚦 Sh | owcard Gothic | | ▼ 9 ▼ | B <i>I</i> <u>U</u> <i>A</i> | | | | | | |

- Click on any text to change its format and use the Formatting Toolbar to do so.
- Double click to rewrite the text itself.
- When you have finished close the Editing window by clicking on the X

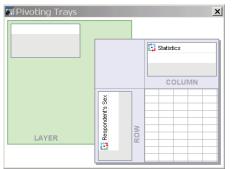


The Formatting Toolbar also gives Pivoting Control (!).



Pivoting control is a useful device, which enables you to change the look of your tables.

Click on the icon to bring up the Pivoting Tray, if it is not already shown.



Clicking on each of the icons at the edges tells you what they represent.

Here the columns are Statistics, and the Rows are Respondent's Sex.

Drag the Statistics icon on to the Row bar so that the 2 are side by side, to see how the table changes; drag it back before proceeding.

- You can copy Output into Word by clicking on it and using Edit > Copy
- In Word use Edit > Paste.

Exporting your Output to Word

- Output can be exported as a Word RTF file or Text file
- Use **File > Export** and select the appropriate entry under Type.

| Objects to Export All All Occument | | | |
|---|---|-------------------------------|---|
| [ype: | | Options: | |
| Word/RTF (*.doc) | - | Layers in Pivot Tables | Honor Print Layer setting (set in Table |
| HTML (*.htm) Portable Document Format (*.pdf) PowerPoint (*.ppt) Text - Plain (*.txt) Text - UTF8 (*.txt) Text - UTF16 (*.txt) Word/RTF (*.doc) | | Include Footnotes and Caption | Yes |
| None (Graphics only) file Name: | • | | |
| c:\Documents and Settings\srx035\Wy Docur | | - | Browse |

Drawing charts

This requires a chapter to itself but the easy way for simple charts is to use **Analyze > Descriptive Statistics > Frequencies**

Then click the chart button and select an appropriate chart. Try it for Region of the United States and draw a bar chart. The dialogue boxes are shown on the next page.

| 😤 Frequencies | Frequencies: Charts |
|---|---|
| Variable(s): Respondent's Sex [s Race of Respondent General Happiness [Number of Brothers Number of Children [Age of Respondent [Highest Year of Sch Hidhest Year School | Statistics Chart Type Charts Eormat Pie charts Pie charts Pie charts Histograms: With normal curve Chart Values |
| Display frequency tables | Erequencies Percentages Continue Cancel Help |

In the same way try drawing a histogram for Age of Respondent.

Exercise

Do not spend too long doing this – the aim is to show you it is much easier drawing charts using Frequencies!

Try drawing the same 2 charts using the Graphs menu and either the Chart Builder or Legacy Dialogs.

After all that ... To return to the data window click on the **under** icon in the toolbar or click on the tab at the foot of the screen, or use the **Window** menu.

The SPSS Tutorial is an extremely useful feature of SPSS

- Click on Help > Tutorial
- Click on the **Introduction** book and take it from there.

Now take a look at the other very useful help: The Statistics Coach.

Click on **Help > Statistics Coach**.

As an example, follow the default settings, and click **Next** each time.

- Summarize, describe or present data Next
- Data in categories
- Tables and Numbers
 Next

Next

• Counts or percentages by category Finish OK

Moving Around

You will be glad to know that the usual short cut keys work here.

| Home | takes you to the first cell of the row you are in |
|-----------|---|
| End | takes you to the last cell of the row you are in |
| Ctrl Home | takes you to the first cell of your data |
| Ctrl End | takes you to the last cell of your data. |

YOUR CHANCE TO CHANGE THE WORLD

Here at Ericsson we have a deep rooted belief that the innovations we make on a daily basis can have a profound effect on making the world a better place for people, business and society. Join us.

In Germany we are especially looking for graduates as Integration Engineers for

- Radio Access and IP Networks
- IMS and IPTV

We are looking forward to getting your application! To apply and for all current job openings please visit our web page: www.ericsson.com/careers



ericsson. com

2. Entering Data

Introduction

This is a chapter for anyone faced with the long and tedious task of entering data. Spend a little time planning this. Wherever possible use numbers rather than text for answers as you can add labels later.

With questionnaires one usually has a separate column for each question, but if you have a question such as:

"Rate each of the following on a score of 1 to 10 according to importance for the community:

Adequate housing Good schools Cultural facilities Sports facilities."

You will need a separate column for each category.

Data can be entered directly or imported from an existing SPSS file, spreadsheet or text file, and we shall cover each of these.

Opening an existing SPSS file.

Use File > Open > Data

Entering Data directly

Entering numbers and text.

The Data Editor Window looks suspiciously like a spreadsheet, and numbers and text can be entered directly.

Be warned, though it looks like a spreadsheet it does not behave like one. Your charts and output will **not** automatically update if you should change the original data, and you cannot enter formulae directly into a cell, though you can do calculations using a different facility.

- Open a new data sheet. Try **Ctrl n**; this is the shortcut key to open a new file.
- Or use **File > New > Data**
- Try entering some numbers in the first column.
- Type what you want in each cell; press the return key or a cursor key.
- If you make a mistake retype the entry.

- Now try to put some text into the same column.
- Can you? You can type it in but when you press Enter it disappears.
- This is because SPSS has identified the column as a numeric one and won't allow any text.
- Put some names of countries in the next column to the right including Australia.
- What happens? Most probably it is cut short.
- Try entering numbers in this column you can but you will not be able to do any calculations with them as SPSS thinks they are text.
- Your new variables have been given the names VAR00001 and VAR00002 which we will now change.

Defining Variables

| Data View Va | ariable View 🛓 | |
|--------------|----------------|--|

At the foot of the screen are two tabs. Click on Variable View to get the following screen.

| Eile Edit | View | Data | Iranstorm | Analyze | Graphs | Unities | Add-ons | Window | Help | | | | |
|-----------|----------|-------|--|---------|--------|------------|-----------|--------|--------|---------|---------|---------|-----------|
| > 🛛 🕰 | <u>⊡</u> | - | 1. III I | 7 # | 1 | ₩ @ | II | 00 | | | | | |
| | | Name | Ту | pe | Width | Decima | ls | Label | Values | Missing | Columns | Align | Measure |
| 1 | VAR | 00001 | Numerio | c C | 1 | 2 | 1 | | None | None | 0 | 冨 Right | 🔗 Scale |
| 2 | VAR | 00002 | String | ε | | 0 | | | None | None | 7 | ≣ Left | 🚓 Nominal |
| 2 | | | and a state of the | | | | | | | | | | |

Overtype VAR00001 and VAR00002 with the names of your new variables: **numbers** and **countries** will do.

Click in the cell under Type to get a grey square.

Click on that to bring up a **Variable Type** box which you can use to define your variable, control the number of decimal places shown, column width etc.

| 🔄 Variable Type 🛛 🗙 |
|-----------------------------------|
| <u>N</u> umeric |
| O_ <u>C</u> omma <u>₩i</u> dth: 8 |
| O Dot Decimal Places: 2 |
| ○ Scientific notation |
| ◯ D <u>a</u> te |
| 🔿 Doļļar |
| O Custom currency |
| ◯ String |
| OK Cancel Help |

Adjusting the width

You can adjust the width of your countries column to 18.

Annoyingly when you return to Data View you will still not find Australia displayed, though when you type it in again it will appear.

Variable names

- They must start with a letter but can now be 64bytes long.
- They can contain numerals e.g. abc12
- But cannot contain spaces or % sign.
- Keep them short.

It is important to keep variable names short so that you can see as much as possible of your data on the screen. It is quite an art to write short names that still give you an idea of what the column is all about. Resist the temptation to write Q1, Q2 etc.

You can enter longer descriptive variable labels to explain what the columns are, and these labels will appear on all output.

Tip: It is better to enter most data as numerical codes and then provide labels explaining what the codes represent. Adding Variable and Value Labels will be explained after you have loaded the spreadsheet.



Entering data via a spreadsheet

Excel spreadsheets can be opened in SPSS with the variable names.

One can also simply copy and paste the data cells from Excel into SPSS but you will have to label the columns.

- To open a spreadsheet use File > Open > Data
- Ask the dialogue box to display **All files** and not just the SPSS ones.

| Look in | : 🕒 My Documents | - 🔁 🍅 🖪 | - |
|-------------|--|-------------|-------|
| 56 | books 📑 Snaøtt Cata | | 117 |
| | | | |
| | Camtasia Studio 🛛 🛅 Snagit Stam | ips | |
| Recent | 📑 Inspiration Data 🛛 🛅 Telkstuff | | |
| | 🛛 🔁 Interprint Albums 🛛 🛅 TylrningPoin | ıt | |
| | 🛛 🛅 My Captivate Projects 🛛 📅 1/29MS wor | rkshops.mht | |
| | 📑 My Data Sources 🦷 💁 þefault.rdp | | |
| Desktop | 🗎 🛅 My eBooks 🛛 🧮 spider.sav | | |
| | A My Music | | |
| | My Pictures | | |
| | My Videos | | |
| y Documents | RecordPad | | |
| | SafeNet Sentinel | | |
| | | | |
| | | | |
| ly Computer | File name: | | Open |
| | | | There |
| | Files of type: All Files (*.*) | | Paste |

Find the spreadsheet to open.

| CDCC will recognize the formet and | 🔛 Opening File Options 🛛 🔀 |
|--|---|
| SPSS will recognise the format and automatically give this dialogue box. | W:VEC\STUDENT\zidney\masters\PULSE4.XLS |
| Tick Read variable names. | Read variable names |
| Click OK. | Range: |
| | Continue Cancel Help |

Adding Variable Labels

To keep your sheet manageable it is advisable to have short column names.

Variable labels can explain more fully the nature of the variable – you have 256 characters for the description.

- In Variable View of the Data Editor.
- Click on the cell under the **Label** column and type in a suitable label.

To give an example I might have the variable **exgrp**, short for exercise group.

The Variable Label for this would then be **exercise group**.

Adding Value Labels

Value Labels explain numerical codes.

To insert a Value Label

- Click in the cell under the Values column and a small grey square appears.
- Click on this to bring up the Define Variable box.

| 1 | | | | | Spelling |
|----------------|------------------------------------|------------------|------------------|------------------|------------------|
| smokes re | gularly | | | | |
| <u>A</u> dd | | | | | |
| <u>C</u> hange |] | | | | |
| <u>R</u> emove |] | | | | |
| | | | | | |
| | smokes re <u>A</u> dd Change | smokes regularly | smokes regularly | smokes regularly | smokes regularly |

Enter a value in the Value box, here it is 1

Type an appropriate label in the Val<u>u</u>e Label box, e.g. *smokes regularly* Click on **Add**

Enter the value 2, and a label, e.g. *non-smoker*

Add

When all the values have been entered use **Add** for the final value, then click on **OK**

A very useful tip for lots of identical value labels for different variables:

- E.g. if you are entering. 0 = No and 1 = Yes,
- Enter them for one variable.
- Then right click on the cell
- Select Copy
- Go to a new variable and use Paste under the Value column.
- This is a huge time saver!

| tie Eat | Yow Data | Iransform Ana | lyze <u>Griebis</u> | Ltitles A | dd-ons Window | FMb. | | | |
|---------|----------|---------------|---------------------|-----------|-------------------|---------------|--------------|---------|---|
| - 🖬 🗛 | | | A 48 | H | *9.0 | | | | |
| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | T |
| 1 | pulse1 | Numeric | 4 | 0 | first pulse rate | None | None | 8 | - |
| 2 | pulse2 | Numeric | 4 | 0 | second pulse rate | None | None | 8 | 4 |
| 3 | ran | Numeric | 4 | 0 | | [1, ran on th | None | 8 | - |
| 4 | smokes | Numeric | 4 | 0 | smoking habits | {1, smokes | <u>Ç</u> opy | | 1 |
| 5 | gender | Numeric | 4 | 0 | | [1, male] | Ejeste | 3 | - |
| 6 | height | Numeric | 6 | 2 | height in inches | None | Grid Font | 3 | 3 |
| 7 | weight | Numeric | 6 | 0 | weight in pounds | None | reone | 8 | 3 |
| 8 | activity | Numeric | 4 | 0 | usual level of p | (1, slight) | None | 8 | 1 |

To return to the data click on the Data View tab at the bottom of the screen.

Important note

When selecting data, defining groups, obtaining multiple response sets you will need to use the numeric value entered in a column and not the text label.

In these circumstances always check what the original data has entered by clicking on the 🔯 icon first.

Finally

It is very easy to make a mistake when entering data.

When it is all entered use **Analyze** > **Descriptive Statistics** > **Frequencies** for each column which will help you spot the most glaring errors .e. 11 instead of 1



Download free ebooks at bookboon.com

21

3. Editing and Handling Data

- Open any SPSS file e.g. 1991 U.S. General Social Survey.sav
- Try each of the following.
- It doesn't matter if you change the data, as long as you don't save the changes.

Correcting entries

Any entry can be over-typed. Click on the cell, type in the correct entry and press Enter. Try changing the value in any cell now.

Deleting entries

- To delete an entry for a cell, click in the cell and press delete.
- Complete columns and rows can be deleted by clicking on the grey cell at the top or side and pressing the **Delete** key on the keyboard.
- Remember the useful Undo icon!

Copying cells, columns and rows

- Cells, columns and rows can be copied by first highlighting them then using the **Edit Copy** menu, or **Ctrl C**.
- Move to where you want them copied and use **Edit > Paste** or **Ctrl V**.

Inserting a variable (a column)

- Click on the top of the column to the right of where you want the new column to appear, i.e. the new column will appear on the left
- Use the Insert Column icon
- **Right Click** at the top of the column to the right of where you want the new column to appear, and use **Insert Variable**

or

• Use Edit > Insert Variable

Inserting a case (a row)

- Click at the side of the row below where you want the new row to appear.
- Use the Insert Row Icon
- **Right Click** at the side of the row below where you want the new row to appear, and use **Insert Cases**.

or

• Edit > Insert Cases

Moving columns

You can drag and drop columns to wherever you like - highlight them first.

Sorting data

SPSS can sort the data, e.g. by Respondent's Sex **Data > Sort Cases**.

| 💶 Sort Cases | × |
|---|---|
| Race of Respondent Region of the United General Happiness [Is Life Exciting or Dull Number of Brothers Number of Children [Age of Respondent [Highest Year of Sch Hidnest Year School Highest Year School | Sort by: Respondent's Sex [sex] Sort Order Ascending Descending |
| OK <u>P</u> aste <u>R</u> eset | Cancel Help |

In the dialog box highlight Respondent's Sex (sex) Click on the arrow to transfer it to the **Sort by** box **OK** Sorting can be Ascending or Descending.

Saving data and output

- Data and output have to be saved **separately**.
- Use **File > Save** in the appropriate window.
- Charts are saved as part of the Output in a .spv file; data is saved as a .sav file.
- You need to save your Output before it can be exported in another format or printed out.
- Be warned Output from SPSS v15 cannot be opened in V16.

Exporting Output

- **Once you have saved** your Output it can also be exported as a Word RTF (Rich Text File) which contains graphics.
- Use **File > Export** and choose Word/RTF from the drop down box.

- Similarly it can exported as a pdf file.
- It is an excellent rule to save frequently.

Saving Data as an Excel file

- SPSS data can be saved as an Excel File.
- Use **Save Data As** and from the drop down box select the appropriate Excel format.
- There are a wide variety of other formats to choose from including csv, dat and SAS.

Copying tables and charts into Word

- In the Viewer window click on what you want to transfer to Word, either a table or chart.
- Use **Edit** > **Copy** and in Word use **<u>Edit</u> > <u>Paste**</u>, or **Ctrl c** and **Ctrl V**.

Printing from SPSS

- Remember that you need to save your Output first.
- You can print directly from the Viewer window using File > Print, but
- use Print Preview first to make sure you have what you want.
- To print just one specific thing click on it first to select it.
- Output that you don't want can be hidden by clicking on the icons in the left hand pane.

Budget-Friendly. Knowledge-Rich.

The Agilent InfiniiVision X-Series and 1000 Series offer affordable oscilloscopes for your labs. Plus resources such as lab guides, experiments, and more, to help enrich your curriculum and make your job easier.



Scan for free Agilent iPhone Apps or visit qrs.ly/po20pli

© Agilent Technologies, Inc. 2012





See what Agilent can do for you. www.agilent.com/find/EducationKit

u.s. 1-800-829-4444 canada: 1-877-894-4414

Anticipate ____Accelerate ____Achieve



Agilent Technologies

Recoding into groups

- You will find it very useful to be able to recode data.
- The **1991 U.S. General Social Survey.sav** data includes the number of brothers and sisters each respondent has in the column headed **siblings**.
- Use Analyze> Descriptive Statistics >Frequencies to get an idea of what this data looks like.

| | | Number of | Brothers a | nd Sisters | |
|---------|-------|-----------|------------|---------------|-----------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 0 | 74 | 4.9 | 4.9 | 4.9 |
| | 1 | 236 | 15.6 | 15.7 | 20.6 |
| | 2 | 276 | 18.2 | 18.3 | 38.9 |
| | 3 | 236 | 15.6 | 15.7 | 54.6 |
| | 4 | 209 | 13.8 | 13.9 | 68.5 |
| | 5 | 118 | 7.8 | 7.8 | 76.3 |
| | 6 | 80 | 5.3 | 5.3 | 81.7 |
| | 7 | 81 | 5.3 | 5.4 | 87.0 |
| | 8 | 58 | 3.8 | 3.9 | 90.9 |
| | 9 | 47 | 3.1 | 3.1 | 94.0 |
| | 10 | 34 | 2.2 | 2.3 | 96.3 |
| | 11 | 22 | 1.5 | 1.5 | 97.7 |
| | 12 | 11 | .7 | .7 | 98.5 |
| | 13 | 9 | .6 | .6 | 99.1 |
| | 14 | 5 | .3 | .3 | 99.4 |
| | 15 | 3 | .2 | .2 | 99.6 |
| | 16 | 1 | .1 | .1 | 99.7 |
| | 17 | 2 | .1 | .1 | 99.8 |
| | 18 | 1 | .1 | .1 | 99.9 |
| | 21 | 1 | .1 | .1 | 99.9 |
| | 26 | 1 | .1 | .1 | 100.0 |
| | Total | 1505 | 99.2 | 100.0 | |
| Missing | DK | 4 | .3 | | |
| | NA | 8 | .5 | | |
| | Total | 12 | .8 | | |
| Total | | 1517 | 100.0 | | |

• It might be useful to regroup the data into subgroups and give each group a numerical code.

| ٠ | As an example I suggest recoding the stude | ents into 3 groups: |
|---|--|---------------------|
| | Those with no brothers or sisters | Group 1 |
| | Those with 1, 2 or 3 brothers or sisters | Group 2 |
| | Those with 4 or more brothers or sisters | Group 3 |

| Use | |
|---|---|
| Transform | Recode into Different Variables |
| Recode Into Different Variables | Respondert's Sex (s Respondert's Sex (s) R |
| Place <i>Number of brothers and Sisters</i> (<i>sibs</i>) in the large box. | Concret Hoppinets [Concret Hoppinets [A lis Life Schlig or OAL Coorge A ge of Respondent [A hope the Structure of Sch |
| Name the new variable <i>sibgrp</i> in the right hand box. | Pripriest Year School Highest Year School Highest Year School Highest Year School Kis Occupational Pre Occupational Pre Cocupational Catego K (costional case selection condition) |
| Click on Change | CK Paste Beset Cancel Help |
| Type in the Label sibling groups | |
| Click on Old and New Values to get the next | |
| dialogue box: | |
| | On the left hand side under Old Value |
| | Click next to Value and enter 0 in the box. |
| O System- or user-missing Old> New: | On the right hand side, as shown |
| C Ragge: | Type 1 in the Value box |
| Range, LOWEST through value: | Click Add. |

Recode the other groups as follows:

Range, value through HIGHEST:

All other values

| Group 2 1, 2, or 3 b For Old Value use | Range 1 through 3 | and for the new Value | 2 |
|--|---|-----------------------|----|
| Don't forget to click of | on Add | | |
| | | | |
| Group 3 4 or more | brothers or sisters | | |
| Group 3 4 or more For Old Value use | brothers or sisters Range, value through highest 4 | and for the new Valu | IP |

Having completed the recoding use Continue OK

You should now have a new column on the right of your data sheet headed sibgrp

<u>W</u>idth:

Output variables are strings

Continue Cancel Help

Revision exercise

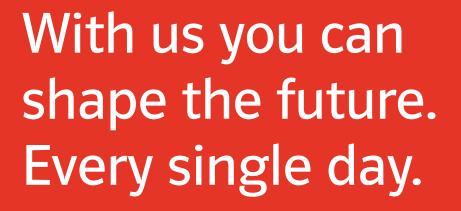
• Provide **labels for** the new variable **sibgrp** to explain what the numbers represent.

Doing Calculations on Variables

Calculations can easily be done in SPSS using Transform > Compute Variable

As an example, in the data **1991 U.S. General Social Survey.sav**, we shall calculate a new column to measure age in months.

- Use **Transform > Compute Variable**
- fill out the dialogue box as shown then **OK**



For more information go to: www.eon-career.com

Your energy shapes the future.



A new variable **agemonths** has been created.

| The age in ye | ars h | nas b | een | m | ulti | plie | ed (*) | by 12 | |
|--------------------------|-------|---------|--------|---------|----------|----------|--------|----------------------|------|
| | | | | | | / | | | |
| Compute Variable | ; | | | | / | | | | |
| Target Variable: | | Numer | IC EXD | ression | | | | | |
| agemonths | = | age * 1 | | V | | | | | |
| Type & Label | | | | | | | | | |
| 💫 Respondent's Sex [s 🔺 | 4 | | | | | | | | |
| Race of Respondent | - | | | | | | | Function group: | |
| Region of the United | | | ~ | | 7 | 8 | 9 | All | - |
| Is Life Exciting or Dull | | | | | <u> </u> | <u> </u> | | Arithmetic | 1000 |
| Number of Brothers | | | <= | >= | 4 | 5 | 6 | CDF & Noncentral CDF | |
| Number of Children [| | | | | 1 | 2 | 3 | Conversion | |
| Age of Respondent [| | | - | | - | - | | Current Date/Time | |
| A thehead Manual Cale | | 1 | 8 | | | 0 | | Date Arithmetic | |

Type in the Numeric Expression 'long-hand' or use the keypad.

The list of functions can be useful for your calculations.

Selecting a subset

During your investigations you may want to look only at the data for the males, or females.

- SPSS enables us to select just these cases using
- Data > Select Cases > If condition is satisfied (click in the circle next to this)
- Click on the **If** ... button under **If condition is satisfied** (the **If button** will not be available if you have not clicked in the circle)
- Enter the appropriate condition, e.g. the example shows what has to be filled in for selecting males.
- Notice you have to put sex =1 not sex = "males"
- This is because the data entered into SPSS in the sex column was numeric

| Region of the United General Happiness [Is Life Exciting or Dull Number of Brothers | <pre>\$ sex=1 + < > 7 8 9 - <= >= 4 5 6</pre> | Function group: All Arithmetic CDF & Noncentral CDF |
|--|---|--|
| Age of Respondent [Age of Respondent [Highest Year of Sch Highest Year School Highest Year School Highest Year School Ris Focupational Pre Coccupational Catego Ris Federal Income T | * = ~= 1 2 3 / 8 1 0 . ** ~ () Delete | Conversion Current Date/Time Date Arithmetic |
| Take Active Part in To Obey [obey] To Be Well Liked or P To Think for Oneself To Work Hard [work To Work Hard [work | | |

| Continue | select Filter out unselected cases | OK |
|----------------------------------|-------------------------------------|----|
| (Tip: Do not delete the other ca | ses as they will be lost for good.) | |

If you scroll down the data sheet you will notice that the females are crossed out on the left, and are now ignored in any operation. Try a frequency table for Respondent's sex and see what you get.

| To restore all the data use | Data >Select Cases > | All cases | OK |
|------------------------------------|----------------------|-----------|----|
|------------------------------------|----------------------|-----------|----|

Be warned: this is all too easily overlooked when you have been working on only part of the data, and then decide to analyse what you think is the complete data set.

Selecting a Random Sample

This is a two stage process:

- First we set the starting point and type of random number generation.
- Then we select the actual sample.

To select the starting point and type of number generator:

• Use Transform > Random Number Generators

| Random Number Generators |
|--|
| Active Generator |
| Set Active Generator |
| SPSS 12 Compatible |
| O Mersenne Twister |
| Active Generator Initialization |
| Set Starting Point |
| Random |
| Eixed Value |
| Value: 2000000 |
| Current Active Generator: SPSS 12 Compatible |
| The active generator setting applies immediately and to future sessions. |
| OK Paste Reset Cancel Help |

- Select Set Active Generator
- There are two ways in which SPSS version 16 generates random numbers. The current active random number generator is displayed.
- You should use Mersenne Twister unless you want to reproduce results generated in SPSS version 12.
- Select Set Starting Point.
- Choosing **Random** allows a different start point for the random selection each time you enter SPSS.
- Entering a Fixed Value (which can be any number) allows a random selection to be reproduced.

- Try them both in the next example and see what happens. •
- If you do not set a starting point you will get the same random selection each time you enter SPSS. •
- Click OK following your selection. •
- Any settings you make will remain in force for future sessions •

To select the actual sample:

- Use: Data > Select Cases > Random Sample of Cases •
- Click on the Sample button
- Fill out the dialogue box appropriately. •

Suppose you wanted to selected a random sample of 4 from the first 9 cases, the box would be set out as follows:

| 🔁 Select Cases: Random Sample 🛛 🔀 |
|--|
| Sample Size |
| Approximately % of all cases |
| ● Exactly 4 cases from the first 9 cases |
| Continue Cancel Help |



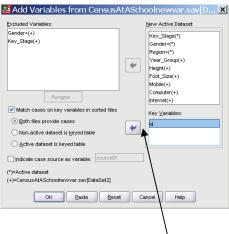
Merging Files

- Sometimes you will have two data files relating to the same people, or two files with similar data but with different people.
- Using Merge Files you can add Variables or Cases to an existing file.

Adding Variables

- Open the first file
- Open the second file which will relate to the same people or objects but with different variables.
- Always open using File Open do not double click from Windows Explorer as this will often open another running of SPSS.
- To merge the two files so that you have all the variables in one:
- You must have a key variable which identifies each case, and you must have sorted the files so that the key variable is in the same order in each.
- Use Data > Merge Files > Add Variables.
- Choose the first file from the list under An open dataset and click Continue

The dialogue box shows an example where the id is the key variable:



Click Match cases on key variables in sorted files and Both files provide cases

Highlight the id in the left hand box and click on the arrow to paste it into the Key Variable box.

Click on OK and OK again at the warning message and the files will merge.

Adding cases

- Open your first file to which you want to add more cases.
- Use Data > Merge Files > Add Cases
- Select An external SPSS data file and click the Browse button, then select your second file.
- Click on **Open**
- There should be no unpaired variables.
- Click on **OK**.
- You should now have a file with all your cases.

Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can neet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering. <u>Visit us at www.skf.com/knowledge</u>

SKF

4. Descriptive Statistics

The Analyze function in SPSS enables us to summarise our data in a number of ways.

The confusion is what to use when, especially as there is often more than one way of doing things in SPSS.

This section provides a guide to what to use, and a brief look at the functions in turn. Remember this is a book on SPSS not on statistics.

Analyze

A 'Very Rough Guide' as to what is appropriate to use when:

All the functions are found under

Descriptive Statistics except where stated.

| Task | SPSS function | Comments |
|---------------------------------------|--|---|
| Counts | Frequencies (offers charts too) Crosstabs | Use %'s as well as counts. %'s are used for comparisons. Round %'s to the nearest whole number in reports. |
| Averages and Measures of spread | Frequencies with the Statistics option; Descriptives. | Make sure you use a sensible measure, e.g. the mean gender is meaningless. |
| Comparing sets of data | Explore (offers charts too) | Beware of using boxplots for inappropriate data, eg nominal. |
| | Crosstabs | Crosstab tables can look untidy, so think carefully about the number of levels and the information required in them. |
| | Analyze > Custom Tables | Use for multiple responses. All tables can be modified. |
| Looking for relationships | Tables: Crosstabs Scatterplots (see Scatter/Dot in the Graphs menu) | Plots and tables give a visual impression of possible relationships: the eyeball test. You may then need to follow this up with the appropriate statistical test. |

The Functions

What follows is a brief description of the following functions:

Frequencies, Descriptives, Explore, Crosstabs and a brief look at other Tables.

Frequencies: Analyze > Descriptive Statistics > Frequencies

This is the best function for overall summaries

Frequencies are used when you want to know how many of something you have.

However, additional statistics available via the **Statistics** button makes it **far more useful** than just counting.

The Charts button is particularly useful; automatically producing charts of your data.

The **Statistics** button brings up the following dialogue box:

| Frequencies: Statistics | × |
|---------------------------------|----------------------------|
| Percentile Values | Central Tendency |
| Quartiles | <u>M</u> ean |
| Cut points for: 10 equal groups | Me <u>d</u> ian |
| Percentile(s): | 🗖 Mode |
| Add | Sum |
| Change | |
| Remove | |
| | Vaļues are group midpoints |
| Dispersion | Distribution |
| Std. deviation | Ske <u>w</u> ness |
| | <u>K</u> urtosis |
| Range S.E. mean | |
| Continue Cancel | Help |

These statistics would be helpful for age but don't be tempted to use them for gender!

Example:

Using the 1991 U.S.General Social Survey.sav data

- Use Frequencies to find the summary statistics for age.
- Draw a histogram of the data.
- Start with Analyze > Descriptive Statistics > Frequencies
- Fill out the dialogue box as shown.
- Click on the **Statistics** button

| D Z D Z D Z D Z Z D Z Z Z Z Z Z Z Z Z Z | |
|---|--------|
| Race of Respondent | Charts |
| Region of the United | Eormat |
| General Happiness [Is Life Exciting or Dull | |
| Number of Brothers | |
| Number of Children [| |
| Highest Year of Sch | |
| Highest Year School | |

- We can ask for the Mean, Median, Std deviation, Minimum and Maximum
- Click **Continue**
- Click on the **Chart** button
- Select Histograms
- Continue OK

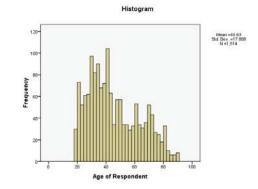


Download free ebooks at bookboon.com

| Frequencies: Statistics | × | 🚰 Frequencies: Charts 🛛 🗶 | |
|--|---|---|--|
| Percentile Values | Central Tendency Mean Median Mode Sum Values are group midpoints | Chart Type None Bar charts Pie charts Histograms: | |
| Dispersion ✓ Std. deviation ✓ Mijnimum ↓ ⊈ariance ✓ Maximum Range SE. mean Continue Cancel | Distribution Skewness Kurtosis Help | Chart Values Erequencies Percentages Continue Cancel Help | |

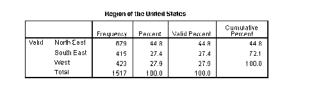
- The Output should look like this.
- No-one would pretend that the histogram is well formatted at this stage but that can be corrected. (See the chapter on charts).
- Believe me, it is by far the quickest way to draw a histogram of age.

| Statistics | | | |
|--------------------|----------------|--------|--|
| _Age of Respondent | | | |
| | N Valid | 1514 | |
| | Missing | 3 | |
| ≯ | Mean | 45.63 | |
| | Median | 41.00 | |
| | Std. Deviation | 17.808 | |
| | Minimum | 18 | |
| | Maximum | 89 | |
| | | | |



Exercise:

Use Frequencies to find the % of respondents living in each of the different regions. Draw a % bar chart to represent this.





Finding Frequencies for Multiple Response Variables

When you write a questionnaire you often include a question where the respondent can tick more than one response.

In the data file 1991 U.S.General Social Survey.sav there are several questions relating to health, e.g.

Are you ill enough to go to a doctor? Have you received counselling for mental problems? Infertility, are you unable to have a baby? Do you have a drinking problem?

Using frequencies we could obtain a separate table for each but SPSS can combine these multiple responses into one table for you.

Use **Analyse > Tables > Multiple Response Sets.**

- First we need to define our Multiple Response set.
- Fill out the dialogue box as shown, with the various health related questions in the Variables box
- Dichotomies Counted value 1 (because there is a 1 in the column when a respondent has that problem)
- Set Name: problems
- Click on Add then OK

| Set Definition Variables in Set: To Obey (obey) To Dink for Oneset To Veik Laked or P To Think for Oneset To Veik Laked (work To Help Others [help Partner (Husband, W Child in hospital [hth7] Child in hospital [hth7] Child in hospital [hth7] Child in tospital [hth7] Child in tospital [hth7] Child in boshild [hth7] Child in b | Mult. Response Sets |
|--|---------------------|
|--|---------------------|

You do not get a table as output but this

Multiple Response Sets

| Name | Coded As | Counted Value | Data Type | Elementary Variables |
|------------|-------------|------------------|-----------|---|
| \$problems | Dichotomies | 1 | Numeric | III Enough to Go to a Doctor Counselling for Mental Problems Infertility, Unable to Have a Baby Drinking Problem |

- Now use Analyze > Tables > Custom Tables
- Your variable **problems** should now appear at the bottom of the Table dialogue box.
- Place it in the Rows and Click **OK**.

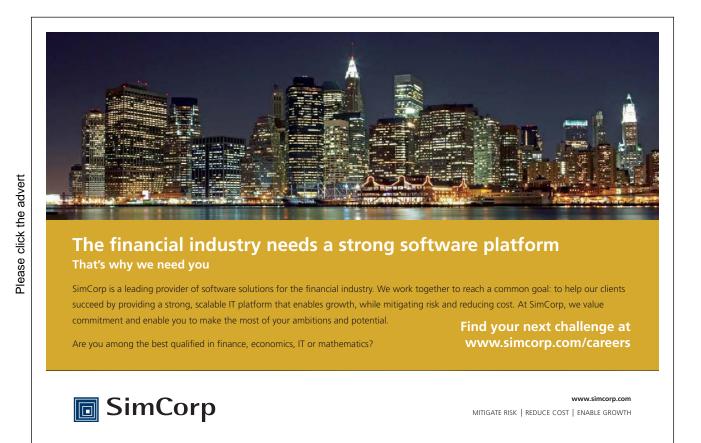
| /ariables: | | | | 1 Normal | E Compact | Layers |
|---|------|----------|--|---|-----------|-------------------|
| Beirg Passed Ove Hevring Trobber W Hoving Trobber W Over Subsets Lo Partner (Husband, Oren's Spouse Bel Most Important Pro Description Proceedings Description Proceedings Description Proceedings Most Important Proce | Rows | \$proble | II Enough Counsel Infertility, Drinking | Count - nonn - nonn - nonn - nonn | guene | |
| Contacting Problem Contacting Proble | | Position | Constants | 5 | • Ude | Category Position |

You should get:

| | Table 1 | |
|------------|---------------------------------------|-------|
| | | Count |
| \$problems | III Enough to Go to a Doctor | 559 |
| | Counselling for Mental Problems | 58 |
| | Infertility, Unable to Have a Baby | 35 |
| | Drinking Problem | 17 |

- For percentages use **the N_%Summary Statistics** button.
- Use Column N%
- Take out the counts by highlighting them and using the back arrow.
- Apply to Selection > OK

| 🚰 Summary Stat | tistics: | | | | | |
|--------------------|----------|------------|------------|--------|----------|----------|
| Selected Variable: | | | | | | |
| Statistics: | 🖌 | Display: | | | | |
| Unweighted Count | | Statistics | Label | Format | Decimals | |
| Responses | | Count | Count | nnnn | 0 | † |
| Row N % | | Column N % | Column N % | nnnn% | 0 | |
| Table N % | | | | | | |
| Subtable N % | • | | | | | |



This should give you:

| | | Table 1 | |
|--------|------|---------------------------------------|------------|
| | | | Column N % |
| \$prob | lems | III Enough to Go to a Doctor | 96% |
| | | Counselling for Mental Problems | 10% |
| | | Infertility, Unable to Have a Baby | 6% |
| | | Drinking Problem | 3% |

Descriptives: Analyze > Descriptive Statistics > Descriptives

Analyze > Descriptive Statistics > Descriptives

Click on **Options** This brings up the following dialogue box:

Descriptives offers much less than Frequencies - only giving a mean for averages, and the standard deviation and range for spread.

| 🖬 Descriptives: Options 🛛 💌 |
|-----------------------------|
| Mean Sum |
| Dispersion |
| Std. deviation V Minimum |
| Ma⊻imum |
| Range S.E. mean |
| Distribution |
| Kurtosis Ske <u>w</u> ness |
| Display Order |
| ④ Variable list |
| <u>○ A</u> lphabetic |
| O Asgending means |
| O Descending means |
| Continue Cancel Help |

Explore: Analyze > Descriptive Statistics > Explore

This is an extremely useful command when you need to compare two sets of data, e.g. ages of males and females. It explores the differences.

The example shows the dialogue box set up to compare the ages of men and women in **the 1991** U.S.General Social Survey.sav data file.

SPSS has been asked to display both statistics and charts, the latter being boxplots and stem and leaf plots - again a very useful automatic facility.

| | Dependent List: | Statistics |
|--------------------------|-------------------------|------------|
| Race of Respondent | Age of Respondent [age] | Plots |
| General Happiness [| | Options |
| Is Life Exciting or Dull | Eactor List: | <u></u> |
| Number of Brothers | Respondent's Sex [sex] | |
| Number of Children [| | |
| Highest Year of Sch | | |
| Highest Year School | Label Cases by: | |
| Highest Year School 💌 | | |
| isplay- | | |
| | | |
| Both OStatistics OP | ots | |

Boxplots are a useful way of comparing two or more data sets. They are as the name implies a box whose length represents the inter-quartile range of the data.

The lower edge of the box is at the lower quartile of the data, and the upper edge at the upper quartile. A horizontal line indicates the median.

'Whiskers' are drawn to the minimum and maximum values within 1.5 box-lengths of each end of the box. Outliers are indicated by o. Values outside 3 box-lengths are indicated by *

Crosstabs: Analyze > Descriptive Statistics > Crosstabs

If you want a table use Crosstabs.

The Tables function is in my opinion only for advanced users of SPSS.

The Crosstabs function produces slightly complex tables, but these can be edited to look tidier.

It has the useful additional facility of doing a Chi-Squared test (and others) if asked - use the **Statistics** button. The **Cells** button enables one to choose Column %'s, Row %'s and Total %'s, but it is advisable to ask for only one at a time, for clarity.

- Using the 1991 U.S.General Social Survey.sav data file.
- The example shows the dialogue box set up to produce a table of **General Happiness** by **Respondent's Sex**.

| 💁 Crosstabs | | x |
|---|--------------------------------|--|
| Race of Respondent [ra Region of the United Sta Is Life Exciting or Dull [life] Number of Brothers and Number of Children [chil | Column(s): | E <u>x</u> act Statistics C <u>e</u> lls Eormat |
| Highest Year of School | -Layer 1 of 1 Previous Next | |
| Display clustered <u>b</u> ar charts | te <u>R</u> eset Cancel Help | |



Download free ebooks at bookboon.com

42

Which gives:

| Count | | | | |
|-------------------|---------------|------------------|--------|-------|
| | | Respondent's Sex | | |
| | | Male | Female | Total |
| General Happiness | Very Happy | 206 | 261 | 467 |
| | Pretty Happy | 374 | 498 | 872 |
| | Not Too Happy | 53 | 112 | 165 |
| | Total | 633 | 871 | 1504 |

General Happiness * Respondent's Sex Crosstabulation

It would be more helpful to give column %'s here to compare the relative happiness of men and women.

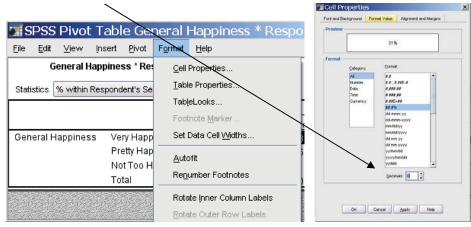
• To do this click on the **Cells** button:

| And fill out the box as shown: | To give | | | | |
|---|-------------------|-----------------------------|--------------|--------------|--------|
| Counts Qbserved Expected | General Hap | piness * Responde 's Sex | ent's Sex Cr | osstabulatio | n |
| Percentages Residuals | | | Re | spondent's S | lex |
| | | | Male | Female | Total |
| <u>R</u> ow <u>Unstandardized</u> <u>Standardized</u> | General Happiness | Very Happy | 32.5% | 30.0% | 31.1% |
| Total Adjusted standardized | | Pretty Happy | 59.1% | 57.2% | 58.0% |
| | | Not Too Happy | 8.4% | 12.9% | 11.0% |
| Noninteger Weights | | Total | 100.0% | 100.0% | 100.0% |
| Round cell counts Round case weights Truncate cell counts Truncate case weights No adjustments Continue Cancel Help | | | | | |

- This table needs formatting to give the %'s as whole numbers.
- Double click on the table To bring up the Pivot Table box

| <u>F</u> ile <u>E</u> dit <u>V</u> iew In | sert <u>P</u> ivot F <u>o</u> rma | t <u>H</u> elp | | | 10000000 |
|---|-----------------------------------|----------------|--------------|--------|----------|
| General Hap | piness * Responde | ent's Sex Cr | osstabulatio | n | |
| Statistics % within Res | spondent's Sex 🔻 | | | | |
| | | Re | spondent's S | Sex | |
| | | Male | Female | Total | |
| General Happiness | Very Happy | 32.5% | 30.0% | 31.1% | |
| | Pretty Happy | 59.1% | 57.2% | 58.0% | |
| | Not Too Happy | 8.4% | 12.9% | 11.0% | |
| | Total | 100.0% | 100.0% | 100.0% | 122 |
| ncence ncenc | ***** | | | | |

- Highlight the cells in the table
- Click on Format > Cell Properties
- Under the **Format Value** tab
- Change **Decimals** to **0**.



NB Producing a table with a variable taking many different values, e.g. age, is not a good idea.

Tables are tricky!

Look at these 2 tables and answer the following questions:

| | | Male | Female | Total |
|-------------------|---------------|------|--------|-------|
| General Happiness | Very Happy | 44% | 56% | 100% |
| | Pretty Happy | 43% | 57% | 100% |
| | Not Too Happy | 32% | 68% | 100% |
| | Total | 42% | 58% | 100% |

| | | Male | Female | Total |
|-------------------|---------------|------|--------|-------|
| General Happiness | Very Happy | 33% | 30% | 31% |
| | Pretty Happy | 59% | 57% | 58% |
| | Not Too Happy | 8% | 13% | 11% |
| | Total | 100% | 100% | 100% |

- What % of females were very happy?
- Of those who were very happy, what % were female?
- The answers are 30% of females were very happy and 56% of those who were very happy were female.
- You may well have got it the wrong way round.
- This is the biggest problem students have wrongly interpreting %'s in tables.

The tip is to do both column and row %'s and have them in front of you so that you can see the difference.

- Crosstabs should produce adequate tables for all your needs, but there are other Tables functions • in SPSS.
- My advice is to ignore these unless you feel very confident.
- Plenty of help on Tables is available under the SPSS Help function.



Go to: Redstarresume.com

5. Charts

Introduction

SPSS provides a wide variety of charts to choose from including bar charts, histograms, pie charts, scatterplots, and boxplots. These are accessed via **Graphs> Chart Builder** Or by Charts > Legacy Dialogs

Charts should convey a message; They should help the reader to understand the data, and not confuse.

Try to use as little 'ink' as possible - cluttered charts are not easy to understand.

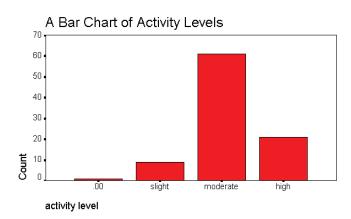
Drawing appropriate charts is not as easy as it looks, so if you feel daunted use the **Charts** options under **Frequencies**.

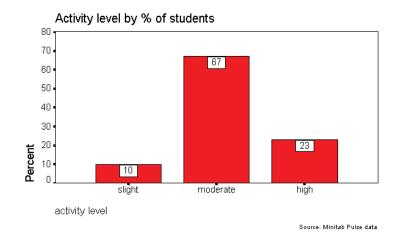
For boxplots use Explore. These two commands will do most of the thinking for you.

In general there are a 2 simple rules which will help:

Decide what your message is and find a chart that conveys it clearly. Label everything, but don't swamp the chart with words - adjust the font size.

Here are 2 examples

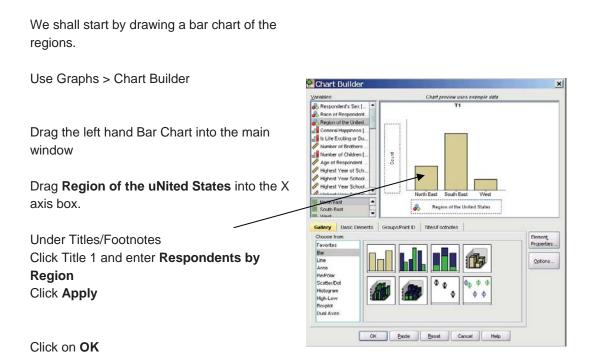




Spot the differences and decide which is more helpful.

A Simple Bar Chart

Using the 1991 U.S.General Social Survey.sav data



Respondents by Region You should get this. To edit the chart double click on it. Count The Chart Editor appears. Depending where you double click on the chart a Properties box should appear with different tabs. North East South East Region of the United States Double click on a bar:

- •
- Under the Fill & Border tab •
- Change the colour;
- Apply. •
- To change the colour of a single bar • click once on one bar;
- it alone will be selected.
- Double click on it
- apply colour as before. •
- Under Depth and Angle do NOT be • tempted to apply shadow or 3-D.
- To remove a category •
- Double click on the x axis labels , e.g. North East
- Click on the Categories Tab
- Highlight North East
- Click on the red cross,
- Apply. •

Percentage Bar Chart

As before use Graphs > Chart Builder

- In the **Element Properties** box (only the top half is shown)
- Select Percentage ()
- Apply
- OK

| Properties | 5 | | × |
|----------------|--|----------------|-------------|
| Depth & Angle | Variables | | |
| Chart Size | Fill & Border | Categories | Bar Options |
| | al level of physica sum) categories k | | % |
| Categories | ouni) outogonoo ii | | |
| Sort by: Val | ue 💌 Direction | : Ascending | • |
| slight | | | |
| moderate | | | |
| a lot | | | |
| | | | × |
| | | | |
| | | | |
| , Excluded: | | | |
| 0 | | | t |
| | | | |
| | | | |
| | | | |
| | | | |
| Lower margin | (%): 5 | Upper margin (| %): 5 |
| | | | |
| | Apply | Cancel | Help |
| | | | |

| 🗜 Element Properties | X |
|-------------------------|---|
| Edit Properties of: | |
| Bar1 | × |
| X-Axis1 (Bar1) | |
| Y-Axis1 (Bar1) | |
| Title 1 | |
| Statistics | |
| Variable: | |
| Statistic: | |
| Percentage () | - |
| Set P <u>a</u> rameters | |

- Double click to edit this chart.
- Click once on a bar to select them all

Then on the Show Data Labels Icon

- Click on Percent
- Transfer to the top box using the arrow ~
- Apply.

Take Count out

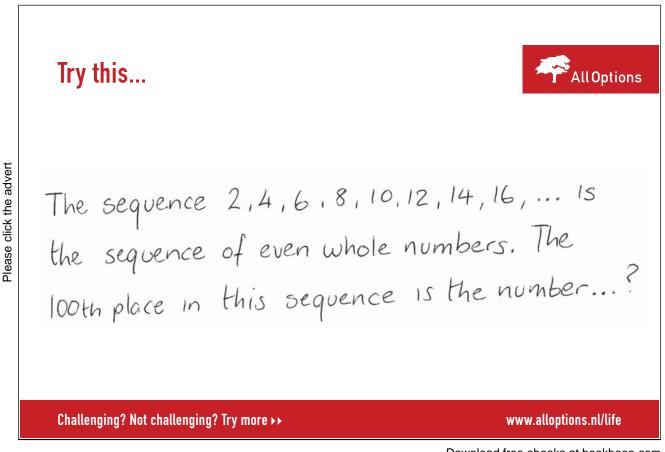
- by highlighting Count, and •
- Using the cross
- Apply.

You can amend the format of numbers by selecting the Number Format tab.

- Always show %'s as whole numbers.
- Type 0 in the Decimal Places box.
- Apply
- Use the Text Style tab to increase the font size: try 12
- Apply.

| Properties | 6 | | | | | | x |
|----------------------|----------|------------------|----------|-------------|----------|---|---|
| Number Format | Da | ta Value La | abels | Variable | s | | |
| Chart Size | Text | Layout | Text | Style | Fill & | Borde | r |
| Labels | | | | | | | 1 |
| Dis <u>p</u> layed: | | | | | | _ | |
| Sount | <u> </u> | <u> </u> | | _ | | ▲ ▼ × | |
| Not Displayed | | | | | - | - | |
| Percent | | nited State: | | s | | t | |
| ○ A <u>u</u> tomatic | : | | | erlapping l | labels | | |
| 🔘 Ma <u>n</u> ual | | 🗹 <u>D</u> ispla | iy conni | ecting line | s to lat | oel | |
| Oustom | | Match | label c | olor to gra | aphic el | lement | |
| | | | | | | | |
| | | Apply | | Close | | <u>l</u> elp | |

÷Г.



To change the %'s on the axis to whole numbers

- Click on the y axis once to select it
- Double click to bring up the properties box
- Select the Number Format tab
- Type 0 in the Decimal Places box
- Click Apply

Transpose the chart using the Transpose icon

The chart can be copied from Output into a Word document using **Edit > Copy** When in Word use **Edit > Paste**.

A clustered bar chart

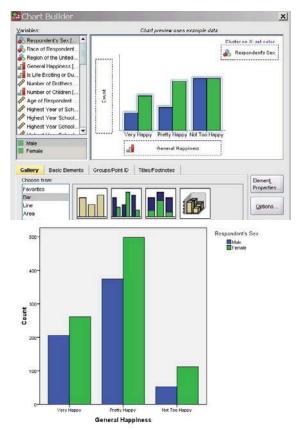
A clustered bar chart is good for comparisons. Here we shall compare the general happiness of males and females.

Use Graphs > Chart Builder Reset

Drag the second bar chart option into the Gallery.

Drag General Happiness into the X axis box Drag Respondent's Sex in to the Cluster on X **box** in the top right of the Gallery window.

You should get



Percentage Clustered Bar Chart

| 🚰 Element Properties 🛛 💌 | t Bullder | × |
|---|---|---|
| Egit Properties of: | Chart preview uses example data | |
| Bart X-Axis1 (Dor1) Y-Axis1 (For 1) GroupColor (Bar1) Statistics Variable Statistic Percentage 0 Set Pgrameters Display gror bars Error Bars Represent | anderst's Serz [| |
| © contidende intervals Level (7) E 35 O Standark error Mutoler: 12 | Beerc Elements Groups/Point D Titles/Footnotes Arres Charges Plements Transpose Element |] |
| For a percentage chart use the Element Pro Bar 1 highlighted Choose Percentage(0) from the Statistic bo Click on Set Parameters | - | |

Choose Total for Each X-Axis Category

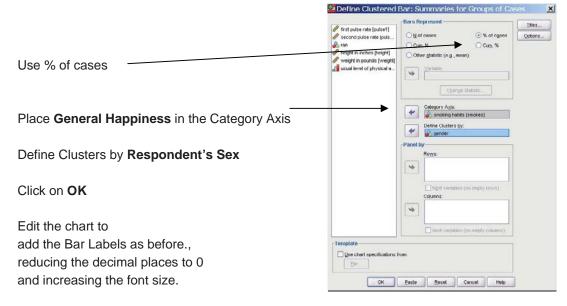
Continue > Apply > OK

Warning: if you apply labels to the bars they will give the wrong %'s.

Percentage Clustered Bar Chart using Legacy Dialogs With correct labels!

For some reason %'s on charts in SPSS pose problems; here is another way of drawing the same chart but with correct labels. It uses the Legacy Dialogs option.

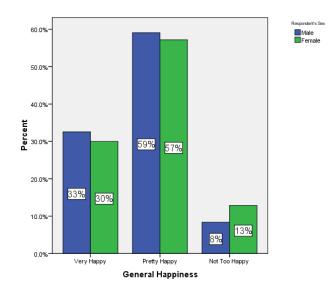
Use Graphs > Legacy Dialogs > Bar... > Clustered Use Summaries for groups of cases > Define

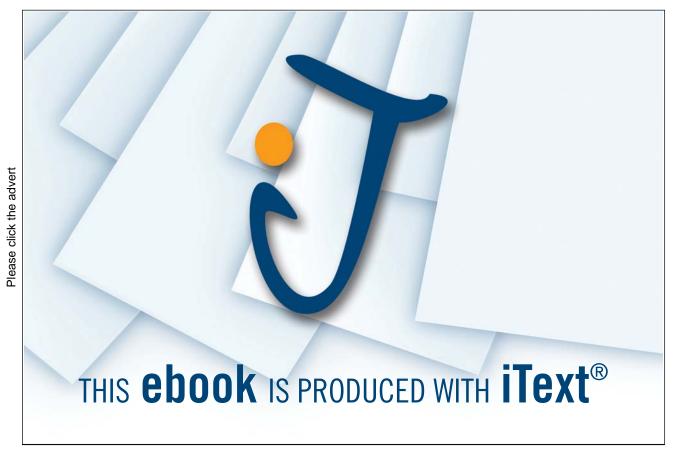


Think of each colour as being a length of ribbon.

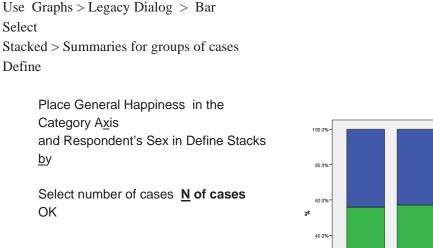
All the ribbons are the same length and represent 100% of each category (males and females).

They are then cut up into the different sections.





BEWARE: If you ask SPSS to add labels to this it will give you the **wrong percentages**. Create a table in cross tabs to find what the %s should be and add the labels as text boxes.



When your chart appears Edit it

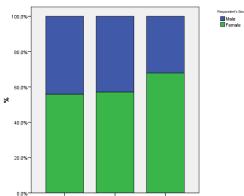
From the menu bar select **Options** At the bottom select **Scale to 100%**

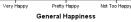
Edit the Y axis label to % by clicking on it. Add text boxes for labels.

Drawing a panel bar chart

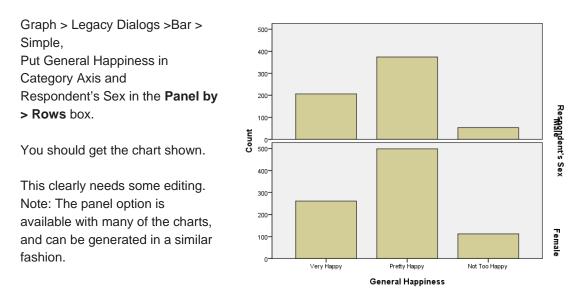
This again uses the Legacy Dialogs.

Panel plots are a style of plot in which subgroups of the data are plotted on separate axes alongside or above and below each other, with the scale on the axes kept common. These can be very useful plots for comparing different subgroups.

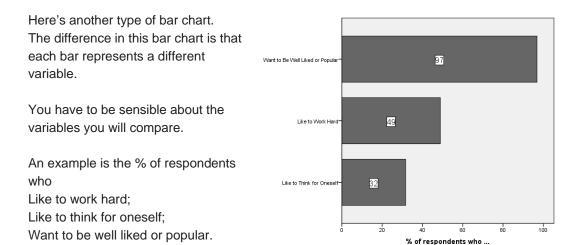




To produce a panel bar chart of physical activity by gender use



Drawing a bar chart of more than one variable



Those who have any of these characteristics are indicated by a 1 in the appropriate column.

| To draw the chart use | 🔁 Bar Charts 🔀 |
|--|--|
| Graphs >Legacy Dialogs > Bar Simple | Simple Clustered Stacked |
| Summaries of separate variables | Data in Chart Are Summaries for groups of cases Summaries of separate variables Values of individual cases Define Cancel |
| At the next dialogue box place each of the | Define Simple Bar: Summaries of Separate Variables |
| activities in the Bars Represent box. | Bars Represent Intex. |
| They will show MEAN(which we will need to change. | Region of the United Region of t |
| Highlight them all by holding down Ctrl and clicking on each. | Highest Year School . Highest Year School . Highest Year School . Bet Scoupational Pre. Occupational Creage Regular School . Nest Year School . Nes |
| Click on Change Statistic. | Commission of the commission o |
| | CK Paste Beset Cancel Help |
| We shall ask SPSS to calculate the % of | 🚰 Statistic 🔀 |
| the entries for each variable less than 2 | Statistic for Selected Variable(s) |
| (since 1 < 2) | Mean of values Standard deviation Median of values ⊻ariance |
| Ask for Percentage below | Moge of values Minimum value Number of cases Maximum value |
| Type 2 in the Value box, | Sum of values Cumulative sum |
| i.e. the % of the numbers in the column < 2 | |
| If we had only wanted a count we would have asked for Number below. | Percentage above Number <u>above</u> Percentile |
| Continue OK | Percentage inside Number inside Values are grouped midpoints Continue Cancel Help |

The resulting bar chart doesn't look quite like the one shown earlier.

By double clicking on it you can open up the Chart Editor window and make the necessary alterations.

You can change the order of the bars. First write down the order you want.

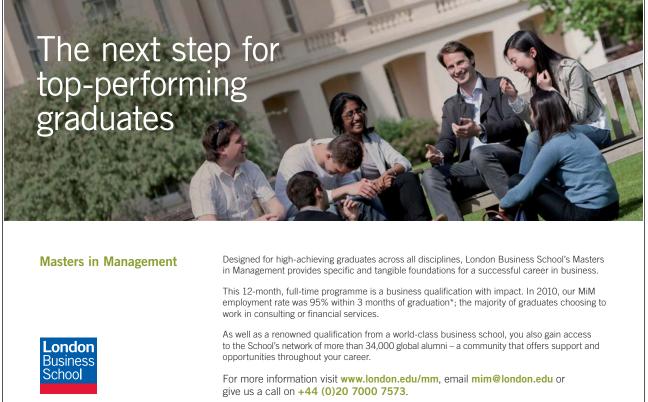
Double click on a bar to open up the **Properties** dialogue box Select the **Categories** tab Highlight the item you want to move under <u>Order</u> Click the up or down arrow. **Apply OK**

| Properties |
|--|
| Depth & Angle Variables Chart Size Fill & Border Categories Bar Options |
| Variable: Variables ▼ Collapse (sun) categories less than: 5 % Categories Sort by: Custom ▼ Direction: Ascending ▼ Order: Want to Be Well Liked or Popular Like to Think for Oneself Like to Work Hard ▼ |
| Excluded: |
| Lower margin (%): 5 Upper margin (%): 5 |

Drawing a pie chart

Please click the advert

Pie charts are used to examine parts of a whole. As an example of drawing one in SPSS we shall draw a pie chart of happiness levels.



* Figures taken from London Business School's Masters in Management 2010 employment report

You can use **Analyze > Descriptive Statistics > Frequencies**... and click on the **Charts...** button to ask for a pie chart,

or use **Graphs > Chart Builder** selecting the Pie Option.

Or use Graphs > Legacy Dialogs >Pie... > Summaries for groups of cases > Define

| Use % of cases | Define Pie: Summaries for Groups of Cases | × |
|--------------------|---|---------|
| | Slices Represent | |
| | Respondent Sev (s Not cases Not case Not ca | Qptions |
| | Age of Respondent [Age to Respondent [Phiphest Year of Sch Phiphest Year School Panel hy | |
| Define Slices by: | Highest Year School Rows: Res Occupational Pre | |
| General Happiness | A Soccupational Pre- Couparing Category R's Federal Income T To Chey Gobey] To Dev (Gobey) To De Well Liked or P To Des Well Liked or P To Think for Courself | |
| Click on OK | To Work Hand [work To Help Otters (Integ.) If III Enough to Go to a | |
| | Template Use chart specifications from: | 1 |

Hopefully you have a similar chart to this.

Open the Chart Editing window by double clicking on the chart.

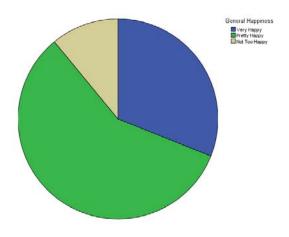
Add labels

To add % to the labels Double click on a slice to bring up the Properties box.

By choosing Percent and General Happiness

and selecting the position of the labels you should be able to get the Pie chart shown on the next page.

See the example dialog box.



| Properties Deta Value Labels Variables Chart Size Text Layout Text Style Fill & Border Labels Disglayed: Percent General Happiness (happy) Vot Displayed: Mot Displayed: | Not Too Happy 31% Very Happy |
|---|------------------------------------|
| Label Position Display Options Automatic Suppress overlapping labels Manual Display connecting lines to label Custom Match label color to graphic element E E E E E E E E E E E E E E E E E E E E E E E E E E E E | |
| | 58% Pretty Happy |

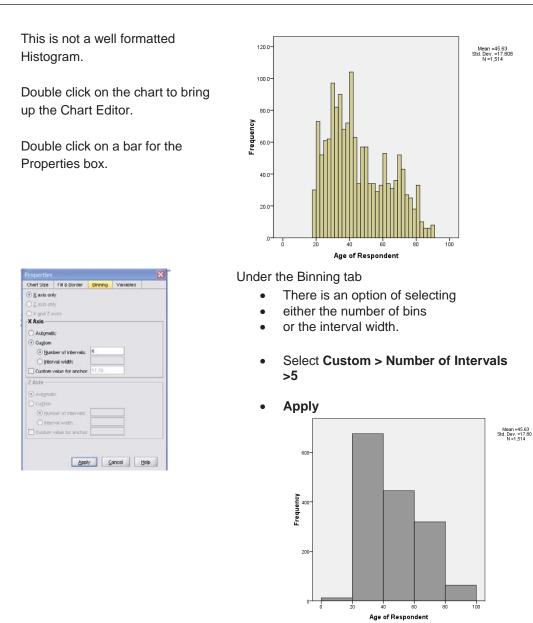
Histogram

Histograms are used for continuous data.

By far the easiest way of drawing a histogram is to use the option under the Chart button in **Analyze** > **Descriptive Statistics >Frequencies**

Alternatively use **Graphs >Chart Builder > Histogram** and drag the first option into the Preview Area. This example shows **Age of Respondent** dragged on to the X axis.

| ariables: | Chart preview uses example data | |
|---|---------------------------------|----------------------------------|
| IS Une Scottig of UU Number of Children I Number of Children I Age of Respondent Highest Vees School Highest Vees School Highest Vees School Re Scottigenonal Pr Occupational Categ Re Scottigenonal Pr No categories (scale variable) | Age of Respondent | |
| Gallery Basic Elements Choose from: Favorites Bar Line Area PsePolar Scatter/Dot Histogram High-Low Boxpick Dual Axes | | Element Properties Qotions |



Should you wish to you can superimpose a variety of curves on the histogram using the distribution curve icon:



Which brings up the following Properties box:

| Properties | | | | |
|------------|--------------------|-----------|--------------------|--|
| Chart Size | Lines | Variables | Distribution Curve | |
| Curves- | | | | |
| | Normal | | | |
| | Uniform | | | |
| | Exponen | tial | | |
| Λ |) <u>P</u> oisson | | | |
| |) <u>O</u> ther cu | rves Beta | * | |

Boxplots

Boxplots are a useful way of comparing two or more data sets.

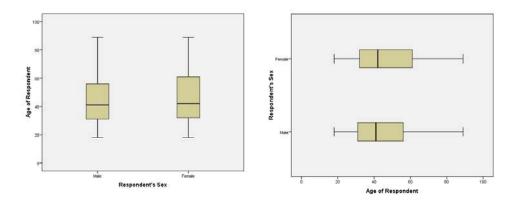
They are as the name implies a box, whose length represents the inter-quartile range of the data.

The lower edge of the box is at the lower quartile of the data, and the upper edge at the upper quartile.

A horizontal line indicates the median.

'Whiskers' are drawn to the minimum and maximum values within 1.5 box-lengths of each end of the box. Outliers are indicated by o. Values outside 3 box-lengths are indicated by * (not shown here).



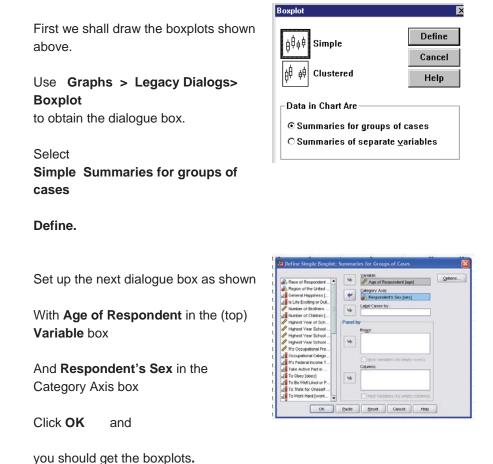


These compare ages by gender.

Boxplots can also be horizontal.

Drawing boxplots can get confusing.

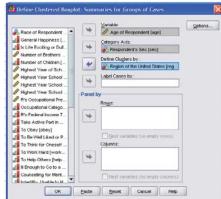
It is easiest to use **Explore** under **Analyze** > **Descriptive Statistics**, but here is how to do it using the **Graphs** menu with two examples to illustrate the differences in different types of boxplots.

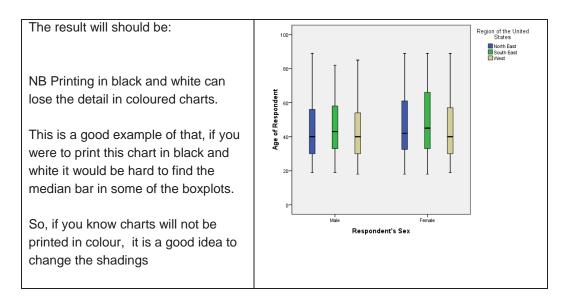


The second example is of a clustered box plot which will show the ages, by gender, for each of the Regions.

Use Graphs > Legacy Dialogs > Boxplot > Clustered Summaries for groups of cases Define.

Complete the dialogue box as follows: **Age of Respondent** in the (top) Variable box **Respondent's Sex** in the Category Axis box **Region of the United** States in the Define Clusters by box. **OK**





There is only one way to master chart drawing in SPSS and that is by having plenty of practice - so over to you.

6. Regression and Correlation

Introduction

In statistics when faced with data we attempt to summarise it and then look for patterns. Regression is about patterns; the possible relationship between **two** sets of data, **bivariate** data.

Open the data set **advert.sav** from SPSS's own sample data sets. This has two columns representing spending on advertising and sales in the same period. The type of questions we might ask about our two variables are:

- Are the two variables related?
- What sort of relationship is there?
- Can we describe (quantify) the strength of the relationship ?
- Can we predict one variable from the other ?



You're full of *energy and ideas*. And that's just what we are looking for.

Looking for a career where your ideas could really make a difference? UBS's Graduate Programme and internships are a chance for you to experience for yourself what it's like to be part of a global team that rewards your input and believes in succeeding together.

Wherever you are in your academic career, make your future a part of ours by visiting www.ubs.com/graduates.

www.ubs.com/graduates

Please click the advert



© UBS 2010. All rights reserved.

Download free ebooks at bookboon.com

63

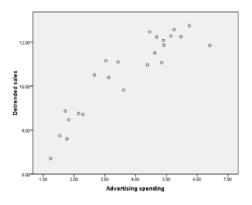
Scatter Diagrams

A visual impression is enormously helpful.

The first thing to do is to plot the data, with the **independent** (x) variable on the **horizontal axis** and the **dependent** (y) variable **vertically.**

Sometimes it isn't obvious which is which. Here it is reasonable to suppose sales depend on advertising.

Plot the data with a Scatter Plot Graphs > Chart builder > Scatter/Dot



Correlation

Correlation quantifies (puts a number to) the strength of the linear relationship between the two variables and also indicates the direction of the relationship.

The correlation coefficient, r, measures the strength of the linear relationship. The value of r is between +1 and -1

Values of **r** close to +1 or -1 represent a strong linear relation. A value of **r** close to **0** means that the linear association is very weak. It could be that there is NO association at all, **or the relationship is non-linear**.

Pearson's product moment correlation coefficient is used where you have variables which represent measurements of some form.

Use **Analyze > Correlate > Bivariate** with the two variables asking for the Pearson coefficient.

| | | Correlations | | |
|---|----------------------|---------------------|----------------------|--------------------|
| | | | Advertising spending | Detrended sales |
| | Advertising spending | Pearson Correlation | 1.000 | .916 |
| | | Sig. (2-tailed) | | .000 |
| ۱ | | N | 24 | 24 |
| | Detrended sales | Pearson Correlation | .916** | 1.000 |
| | | Sig. (2-tailed) | .000 | |
| | | N | 24 | 24 |

**. Correlation is significant at the 0.01 level (2-tailed).

This shows a correlation coefficient of 0.916 and a significance value of 0.000.

The significance is <0.05 and indicates that if there is no linear relationship between spending on advertising and sales there is a less than 0.05% chance that a random sample of this size would give a value of **r** as extreme as 0.916.

Spearman's rank correlation coefficient can also be used. Spearman's coefficient can be used when you have merely ordered variables, e.g. treatments **ranked** as to effectiveness. The printout gives a different value for r having been calculated another way, but the significance value is again <0.05.

| | | Correlations | | |
|----------------|----------------------|-------------------------|-------------------------|--------------------|
| | | | Advertising spending | Detrended sales |
| Spearman's rho | Advertising spending | Correlation Coefficient | 1.000 | .889** |
| | | Sig. (2-tailed) | | .000 |
| | | N | 24 | 24 |
| | Detrended sales | Correlation Coefficient | .889** | 1.000 |
| | | Sig. (2-tailed) | .000 | |
| | | N | 24 | 24 |

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation and Causation

Correlation quantifies the degree of association between two variables- **BUT BEWARE** for although two variables may seem to be related, a change in one may not cause a change in another.

Correlation coefficients **are the most frequently misused statistics** so when interpreting your correlation coefficient remember

- that correlation does not mean causation;
- to use your common sense !

Regression

Having discovered that two variables are correlated the next question might be can we model this data using a straight line?

Can we predict what the sales between and are likely to be from the spending on advertising?

Linear Regression is the technique that is used to find the line that best models the data. We first have to decide which variable is dependent on the other – here the sales are likely to be dependent on the spending on advertising.

Use Analyze > Regression > Linear

Place **Detrended sales** in the Dependent box and **Advertising spending** in the independent box.

| Linear Regression | × |
|-------------------------|---|
| Advertising spending [a | Dependent: Statistics Ø Detrended sales [sales] Plots I of 1 Save yous Next Independent(s): Options |
| OK Paste | Method: Enter |

The output is:

| Variables | Entered/Removed ^b |
|-----------|------------------------------|
|-----------|------------------------------|

| Mode I | Variables Entered | Variables Removed | Method |
|-----------|--------------------------------------|----------------------|--------|
| 1 | Advertising spending ^a | | Enter |
| - 01 | | | |

a. All requested variables entered.

b. Dependent Variable: Detrended sales

Model Summary

| Mode | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|------|-------|----------|----------------------|-------------------------------|
| 1 | .916ª | .839 | .832 | .73875 |
| | | | | |

a. Predictors: (Constant), Advertising spending

 $\textbf{ANOVA}^{\rm b}$

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|-------------------|----|-------------|---------|-------|
| 1 | Regression | 62.514 | 1 | 62.514 | 114.548 | .000ª |
| | Residual | 12.006 | 22 | .546 | | |
| | Total | 74.520 | 23 | | | |

a. Predictors: (Constant), Advertising spending

b. Dependent Variable: Detrended sales

Coefficients^a

| | | Unstandardize | d Coefficients | Standardized Coefficients | | |
|-------|----------------------|---------------|----------------|------------------------------|--------|------|
| Model | | В | Std. Error | Beta | t | Siq. |
| 1 | (Constant) | 6.584 | .402 | | 16.391 | .000 |
| | Advertising spending | 1.071 | .100 | .916 | 10.703 | .000 |

a. Dependent Variable: Detrended sales

The top box showing the variables entered is self explanatory.

The Model Summary shows the goodness of fit statistics indicating whether the model is a good fit.

- **R** is the **correlation coefficient** measuring the strength of the linear relationship.
- **R Square** is the **coefficient of determination**, more usually expressed as a percentage. Here it tells us that 89% of the variability in the sales can be explained by the variability in the spending on advertising.

• The **Std Error of the Estimate** can be thought of as a typical residual; the difference between what is predicted by the model and what is observed.

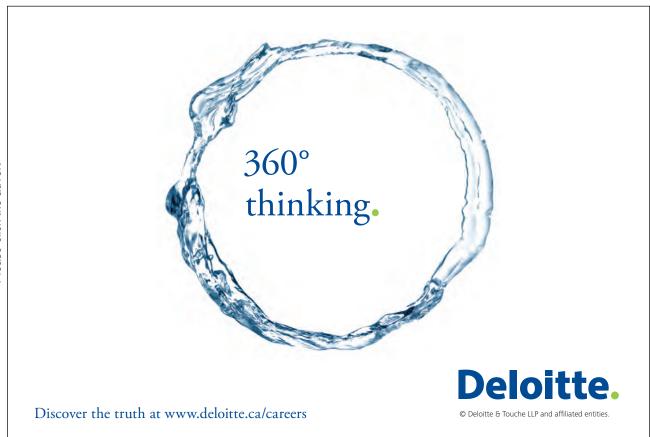
The **ANOVA** box shows a significance value of .000 This indicates that the regression is significant, i.e. that there is a useful linear model.

The **Coefficients** box tells us that the equation that models the line has a **slope of 1.071** and an **intercept of 6.584**.

We need to know if the variable is actually significant. This is indicated by the significance column on the right. Sig values > 0.05 indicate that the coefficient is not significant. Remember that we are trying to deduce a model to predict price for the population based on a relatively small sample. This means our values for the coefficients of the slope and intercept are only **estimates**.

The t value column has done a t-test to test the probability that the coefficient is zero given the sample data, and the Sig column is the p value for this test.

Here our coefficients are OK so our regression equation would be sales = 1.071^* spending on advertising + 6.584



We need to know:

- 1. Is this line a good fit?
 - The answer is given by the goodness of fit statistics. and
- Is it an appropriate model?
 Here we need to look at the residual plots available under the Plots button
- To obtain a chart showing the regression line use **Analyze** > **Regression** > **Curve Estimation**
- filling out the dialogue box as shown.

| 🚰 Curve Estimation | × |
|----------------------------|------|
| | Save |
| OK Paste Reset Cancel Help | |

We have looked at **linear Regression** but there are other models available from the Regression menu.

Multiple Regression

Multiple regression is used where we have more than one variable which might predict the dependent variable.

For a linear model we use the same commands as before: Analyze > Regression > Linear

But place more than one variable in the Independents box

This output gives us the values of the coefficients.

Again, we need to know which variables are actually significant.

This is indicated by the significance column on the right.

Sig values > 0.05 indicate that the coefficient is not significant.

Remember that we are trying to deduce a model to predict price for the population based on a relatively small sample. This means our values for the coefficients are only estimates.

The t value column has done a t-test to test the probability that the population coefficient is zero given the sample data, and the Sig column is the p value for this test.



7. Statistical Tests

Many students and others want to be able to use the statistical tests in SPSS for hypothesis testing. This is not a statistics textbook, but a guide to using SPSS, so no theory is included but it is nevertheless important to stress that you need:

- To be clear about your research question, or the hypothesis you propose to test.
- To be sure that the data you are collecting will actually answer that research question, and
- To collect it from a random sample, to be free from bias.

The procedure is:

- Write your hypothesis and null hypothesis.
- Collect the data.
- Look at the data what does the evidence of the sample suggest?
- Make a chart if possible.
- It is usual to test the Null Hypothesis which is a statement of no difference; no association.
- Select an appropriate test.
- Check that the requirements for that test have been satisfied; e.g. was the sample a random sample?
- Carry out the test and identify the p value.
- Is the p value >= 0.05, or < 0.05?

| Probability | Р | Significance | Decision |
|--------------------------------|---------|---------------------------|------------------------------|
| Less than 1 in 10,000 | < .0001 | Significant at .01% level | Reject null hypothesis |
| Less than 1 in 1000 | < .001 | Significant at .1% level | Reject null hypothesis |
| Less than 1 in 100 | < .01 | Significant at 1% level | Reject null hypothesis |
| Less than 5 in 100 | < .05 | Significant at 5% level | Reject null hypothesis |
| More than or equal to 5 in 100 | >= .05 | Not significant | Don't reject null hypothesis |

Table of P Values and Significance

- Decide if the evidence supports the null hypothesis.
- State the decision about the original hypothesis.

In the examples that follow we shall use the data file 1991 U.S.General Social Survey.sav .

Confidence Intervals: Analyze > Descriptive Statistics > Explore

The requirement for this test is that the sample has been randomly selected.

Use this to test for a hypothesised value; it will give you the confidence interval for the mean of a population.

E.g. Test the hypothesis that the mean number of brothers and sisters people have is 3.

Using Analyze > Descriptive Statistics > Explore

- with Age of Respondent in the Dependent List
- with no Factor
- asking for Statistics only

| | Dependent List: | Statistics |
|-------------------------|------------------------|------------|
| Respondent's Sex [s | Number of Drothers and | - |
| Race of Respondent | 4 | Piots |
| Region of the United | | Options |
| General Happiness [| Eactor List: | |
| Is Life Exciting or Dul | | |
| Number of Children [| | |
| Age of Respondent [| | |
| Highest Year of Sch | Label Cases by: | |
| PHighest Year School 💌 | | |
| Display | | |
| | | |
| Doth 💿 Statistics 🔿 Plo | | |

The output is:

Descriptives

| | | | Statistic | Std. Error |
|------------------------|-------------------------|-------------|-----------|------------|
| Number of Brothers and | Mean | | 3.93 | .079 |
| Sisters | 95% Confidence Interval | Lower Bound | 3.78 | |
| | for Mean | Upper Bound | 4.09 | |
| | 5% Trimmed Mean | , | 3.69 | |
| | Median | / | 3.00 | |
| | Variance | | 9.282 | |
| | Std. Deviation | | 3.047 | |
| | Minimum | | 0 | |
| | Maximum | | 26 | |
| | Range | | 26 | |
| | Interquartile Range | | 3 | |
| | Skewness | | 1.468 | .063 |
| | Kurtosis | / | 3.507 | .128 |

The confidence interval would support any hypothesis which suggested that the population mean was between the Lower Bound of 3.78 and the Upper Bound of 4.09

There is no evidence at the 5% level that the mean number of brothers and sisters is 3.

The One-Sample T test

The requirement for this test is that the sample has been randomly selected.

This is an alternative method to using confidence intervals.

Use this to test for a hypothesised value.

E.g. Test the hypothesis that the mean number of brothers and sisters people have is 3.

Use Analyze > Compare Means > One-Sample T test

| 🔁 One-Sample T Test | × |
|--|---|
| Respondent's Sex [s Region of the United General Happiness [Is Life Exciting or Dull Number of Children [Age of Respondent [| Test Variable(s): Number of Brothers and |
| | Test <u>V</u> alue: 3 |
| OK Paste | Reset Cancel Help |

Place **Number of Brothers and Sisters** in the **Test Variable** box And type 3 in the **Test Value** box

The output is:

| One-Sample Statistics | | | | |
|-----------------------------------|------|------|----------------|--------------------|
| | N | Mean | Std. Deviation | Std. Error Mean |
| Number of Brothers and Sisters | 1505 | 3.93 | 3.047 | .079 |

| Une-Sample Test | | | | | | |
|-----------------------------------|--|------|-----------------|--------------------|-------|-------|
| Test Value = 3 | | | | | | |
| | 95% Confidence Interval of the Difference | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | Lower | Upper |
| Number of Brothers and Sisters | 11.862 | 1504 | .000 | .932 | .78 | 1.09 |
| | | | | | | |

One Comple Test

The significance value is < 0.000 which shows that there is a significant difference between 3 and the mean number of brothers and sisters of those in the sample.

The Chi-Squared Test for contingency tables

The requirements for this test are that the samples are random and at least 80% of the cells in the table should have expected counts of at least 5 and no cell should have an expected count less than 1.

| The question: | Is there an association between happiness and gender? |
|--------------------------|---|
| The Research Hypothesis: | There is an association between happiness and gender. |
| The Null Hypothesis: | There is no association between happiness and gender. |

Use Analyze > Descriptive Statistics > Crosstabs

Complete the dialogue box as shown

| | Row(s): | Exact |
|--|-------------------------------|------------|
| Race of Respondent [ra Region of the United Sta | General Happiness [happy] | Statistics |
| Is Life Exciting or Dull [life] | | Cells |
| Number of Brothers and | <u>C</u> olumn(s): | - |
| Number of Children [chil | Respondent's Sex [sex] | Eormat |
| Highest Year School Co Highest Year School Co Highest Year School Co Highest Year Schoul Co Ro Occupational Preati Occupational Category [R's Federal Income Tax Take Active Part in Worl | Layer 1 of 1 Preylous Uext | |
| Display clustered bar charts | | |
| Territori, conserve Territoria | | |



- Click on the **Statistics** button
- Click in Chi-Squared (top left box)
 Continue

| 🔁 Crosstabs: Statis | tics 🗙 | | | |
|--|-------------------------|--|--|--|
| ✓ Chi-square | Correlations | | | |
| Nominal | Ordinal | | | |
| Contingency coefficient | <u>G</u> amma | | | |
| Phi and Cramer's V | Somers' d | | | |
| Lambda | Kendall's tau- <u>b</u> | | | |
| Uncertainty coefficient | Kendall's tau- <u>c</u> | | | |
| Nominal by Interval | <u>K</u> appa | | | |
| Eta | Risk | | | |
| | McNemar | | | |
| Cochran's and Mantel-Haenszel statistics Test common odds ratio equals: | | | | |
| Continue Cancel | Help | | | |

- Click on the **Cells** button
- for Counts: Observed Expected

| • | Continue |
|---|----------|
| | |

| 🚰 Crosstabs: | Cell Display 🗙 |
|-------------------|--|
| Counts | ٦ |
| Deserved | |
| Expected | |
| Percentages | Residuals |
| Row | Unstandardized |
| Column | Standardized |
| <u> </u> | Adjusted standardized |
| -Noninteger Weig | jhts |
| Round cell cou | nts ORound case <u>w</u> eights |
| 🔘 Truncate ceļi c | ounts 🔘 Truncate case weig <u>h</u> ts |
| O No adjustments | 3 |
| Continue | Cancel Help |

and then on $\boldsymbol{O}\boldsymbol{K}$

This should bring up the following Output. By looking at the table of expected and observed counts one can see that there are more men who are happy than expected and more women who are Not Too Happy (the eyeball test).

| | | | Respondent's Sex Male Female Total | | Sex |
|-------------------|---------------|----------------|---------------------------------------|-------|--------|
| | | | | | Total |
| General Happiness | Very Happy | Count | 206 | 261 | 467 |
| | | Expected Count | 196.5 | 270.5 | 467.0 |
| | Pretty Happy | Count | 374 | 498 | 872 |
| | | Expected Count | 367.0 | 505.0 | 872.0 |
| | Not Too Happy | Count | 53 | 112 | 165 |
| | | Expected Count | 69.4 | 95.6 | 165.0 |
| | Total | Count | 633 | 871 | 1504 |
| | | Expected Count | 633.0 | 871.0 | 1504.0 |

General Happiness * Respondent's Sex Crosstabulation



| | Value | df | Asymp. Sig. (2-sided) | |
|---------------------------------|--------|----|--------------------------|--|
| Pearson Chi-Square | 7.739ª | 2 | .021 | |
| Likelihood Ratio | 7.936 | 2 | .019 | |
| Linear-by-Linear Association | 4.812 | 1 | .028 | |
| N of Valid Cases | 1504 | | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 69.44.

So it comes as no great surprise that the value of Chi-squared (7.739) is significant because the p value is 0.021 The null hypothesis is not accepted.

The conclusion is that this sample shows evidence at the %5 level that there is an association between happiness and gender, with men appearing to be happier.

t-test for related samples

The requirement for this test is that the sample is randomly selected. There is no need for the underlying population to be normal provided the sample size is large, i.e. >30.

With related samples we are comparing the differences between **pairs of readings that are related**: two pulse readings from the same patient.

Use the SPSS data set **New drug.sav** for this example. This is a very small data set but we shall assume the subjects were randomly selected.

| The question: | Is there a difference in the population means of the first and second pulse rates of each patient? |
|--------------------------|---|
| The Research Hypothesis: | There is a difference in the population means of the first and second pulse rates of each patient. |
| The Null Hypothesis: | There is no difference in the population means of the first and second pulse rates of each patient. |

Use Analyze > Compare Means > Paired-Samples T Test

| The dialogue box should be | 🚰 Paired-Samples T Test | x |
|-------------------------------------|---|-----------------------|
| completed by clicking on | Paired <u>Variables</u> : | Options |
| Pulse, Time1 | Second pulse rate [puls 1 Second for the pulse rate [puls | |
| clicking on the arrow | ran smoking habits (smokes) smoker pender / height in inches (height) | * |
| and then on Pulse Time2 | weight in pounds [weight] usual level of physical a | + |
| and on the arrow | | \longleftrightarrow |
| to place them in the variables box. | | |
| | OK Paste Reset Cancel | Help |

οκ

You should obtain the following Output:

| Time 1 Time 2 Paired | 1ean 2.433 2.517 Sample | 12 12 s Correlatio | Std. Deviation .2605 .3326 | Mean .0752 .0960 |] | | | | |
|----------------------------|----------------------------------|--------------------------|----------------------------------|------------------------|---------------------------|---|---|---|---|
| Paired | | | | .0960 | J | | | | |
| | ▲ Sample | s Correlatio | | | | | | | |
| | Sample | s Correlatio | | | | | | | |
| | | | ns | | | | | | |
| | | N | Correlation | Sig. | | | | | |
| Time 1 & Pul: | se, | 12 | .969 | .000 | | | | | |
| | | | | | | | | | |
| | | | | Paired Samples | s Test | | | | |
| | | | | | Paired Differe | ences | | | |
| | | | | | 95% Confidence Differe | Interval of the | | | |
| | | Mean | Std. Deviation | Std. Error Mean | Lower | Upper | t | df | Sig. (2-tailed |
| Time 1 - Puls | e, | 0833 | .1030 | .0297 | 1488 | 0179 | -2.803 | 11 | .01 |
| | Time 1 - Puls | Time 1 - Pulse, | | | Mean Std. Deviation Mean | 95% Confidence Differe Std. Error Mean Std. Deviation Mean Lower | Paired Differences Paired Differences 95% Confidence Interval of the Difference Std. Error Mean Std. Deviation Mean Lower Upper | Paired Differences Paired Differences 95% Confidence Interval of the Difference Std. Error Mean Std. Deviation Mean Lower Upper t | Paired Differences Paired Differences 95% Confidence Interval of the Difference Mean Std. Deviation Mean Lower Upper t df |

By looking at the sample means one can see they are different. The p value is 0.017 showing that the t value is significant.

The null hypothesis is rejected.

The conclusion is that this sample shows there is a significant difference between the population means of the first and second pulse rates of patients.



t-test for the differences in the Means of independent samples

The requirement for this test is that the samples are randomly selected. There is no need for the underlying population to be normal provided the sample sizes are large, i.e. >30.

Here we are comparing the differences between pairs of readings that are not related. We shall use the data file **1991 U.S.General Social Survey.sav**

| The question: | Is there a difference in the highest year of school completed by males and |
|--------------------------|--|
| | females? |
| The Research Hypothesis: | There is a difference in the highest year of school completed by males and |
| | females. |
| The Null Hypothesis: | There is no difference in the highest year of school completed by males |
| | and females |

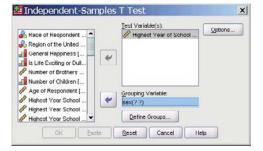
Use Analyze > Compare Means > Independent-Samples t Test

Place Highest Year of School in the Test Variable box and

sex in the Grouping Variable

Click on **Define Groups.**

| Define Groups | | | | |
|----------------------|-----------|------|--|--|
| O Use specifi | ed values | | | |
| Group <u>1</u> : | 1 | | | |
| Group <u>2</u> : | 2 | | | |
| ◯ <u>C</u> ut point: | | | | |
| Continue | Cancel | Help | | |



Fill out the box as shown.

The 1 and 2 are the codes for males and females.

You should get the following Output (which is annoyingly wide).

Group Statistics

| | Respo nden | N | Mean | Std. Deviation | Std. Error Mean |
|------------------------|---------------|-----|-------|----------------|--------------------|
| Highest Year of School | Male | 633 | 13.23 | 3.143 | .125 |
| Completed | Female | 877 | 12.63 | 2.839 | .096 |

| | - | for Equ | s' Test ality of | | | | | | | |
|--|-----------------------------------|---------|---------------------|-------|----------|---------|------------|------------|-------|-------------------------|
| | | | | | | Sig. | | | Confi | 5% dence I of the |
| | | | 1 | | | (2- | Mean | Std. Error | Diffe | rence |
| | | F | Sig. | t | df | tailed) | Difference | Difference | Lower | Upper |
| Highest Year of School Completed | Equal variances assumed | 11.226 | .001 | 3.887 | 1508 | .000 | .602 | .155 | .298 | .906 |
| | Equal variances not assumed | | × | 3.824 | 1276.454 | .000 | .602 | .157 | .293 | .911 |

Independent Samples Test

Using the eyeball test again, looking at the means reveals a difference in the sample means. Levene's test indicates, by the p value, whether we should assume equal or unequal variances. If the p value is < 0.05 the evidence suggests that the variances are unequal.

Here p=0.001 so we use the Equal variances **not assumed** line for the t test for the means.

This gives a low p value of < 0.0005 so we conclude that the samples show that there is a significant difference between the population means of the highest year of school completed by male and females.

Analysis of Variance

We are assuming here that we have independent simple random samples drawn from normal populations.

Analysis of variance is a method for comparing the means of several populations. Simple random samples are drawn from each and are used to test the null hypothesis that the population means are all equal. ANOVA compares the variation among groups with the variation within groups.

| The question: | Is there a difference in the population means of the Highest year of school completed for each region? |
|--------------------------|---|
| The Research Hypothesis: | There a difference in the population means of the Highest year of school completed for each region. |
| The Null Hypothesis: | There is no difference in the population means of the Highest year of school completed for each region. |

• Use Analyze > Compare Means > One-Way ANOVA

Fill out the dialogue box as shown with

the **Highest Year of School** in the **Dependent List**,

and **Region of the United States** as the **Factor**.

| | | Dependent List: | [|
|--------------------------|---|---------------------------|-----------|
| Respondent's Sex [s 🔺 | | Highest Vear of School | Contrasts |
| Race of Respondent | | | Post Hoc |
| Conoral Happiness [| | | Options |
| Is Life Exciting or Dull | * | | |
| Number of Brothers | | | |
| Number of Children [| | | |
| Age of Respondent [| | | |
| Highest Year School | | Eactor: | |
| Highest Year School | * | Region of the United Stat | |

Click on the **Options** button and select **Descriptive Statistics**;

The Output is:

Oneway

Descriptives

| Highest Year of School Completed | | | | | | | | |
|----------------------------------|------|-------|----------------|------------|--------------------|-------------|---------|---------|
| | | | | | 95% Confider Me | | | |
| | N | Mean | Std. Deviation | Std. Error | Lower Bound | Upper Bound | Minimum | Maximum |
| North East | 676 | 13.00 | 2.778 | .107 | 12.79 | 13.21 | 3 | 20 |
| South East | 411 | 12.46 | 3.352 | .165 | 12.13 | 12.78 | 0 | 20 |
| West | 423 | 13.11 | 2.885 | .140 | 12.83 | 13.38 | 3 | 20 |
| Total | 1510 | 12.88 | 2.984 | .077 | 12.73 | 13.03 | 0 | 20 |

| <u>Highest Year of Scl</u> | hool Completed | | | | |
|----------------------------|-------------------|------|-------------|-------|------|
| | Sum of Squares | df | Mean Square | F | Siq. |
| Between Groups | 104.635 | 2 | 52.317 | 5.914 | .003 |
| Within Groups | 13332.084 | 1507 | 8.847 | | |
| Total | 13436.719 | 1509 | | | |

ANOVA

The p value is 0.003 which is <0.05, so we conclude that there is evidence to suggest that that the means of the 3 populations are not all the same.

Non-Parametric Tests

A **parameter** is a number describing the **population**, e.g. the mean or standard deviation, as distinct from a **statistic** which is a number that can be calculated from the **sample** data without needing to know anything else about the population.

Many statistical tests are parametric tests and make the assumption that the populations involved have 'normal distribution'. These tests are very useful and robust but there are occasions when we would like to compare two samples which we cannot assume come from a 'normal' population, or where the measurements are on an ordinal scale as distinct from an interval one.

For such populations we use **non-parametric** tests. We can use these on 'normal' data too.

Note: if the values in the population have a skewed distribution, or if the measurement scale is ordinal then it is better to use the median rather than the mean.

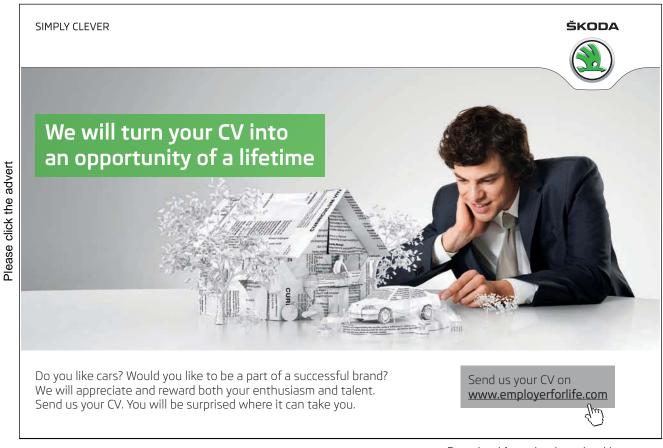
Wilcoxon Rank-Sum Test also known as the Mann Whitney U test for independent samples

| The question: | Is the population median of the Highest Year of School Completed the same for males and females? |
|----------------------|---|
| The Hypothesis: | There is a difference in the population median of the Highest Year of School Completed for males and females? |
| The Null Hypothesis: | There is no difference in the population median of the Highest Year of School Completed for males and females? |

First we need to find the median Highest Year of School Completed for males and females. Use **Explore**. males: 13.23; females: 12.63.

These are two independent samples; the variable (Highest Year) we shall treat as continuous. Use **Analyze > Nonparametric Tests > 2 Independent Samples**

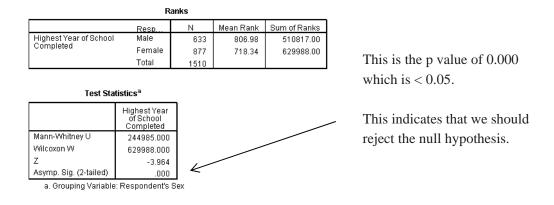
Complete the dialogue box as shown using the **Define groups** button for the genders (1, 2).



80

| Two-Independent-Samples Tests | |
|---|--|
| Rece of Respondent Test Variable List Exact Region of the United Highest Year of School Options General Happiness [Image: Comparison of Brothers Options Number of Brothers Grouping Variable: Sex.(? ?) Highest Year School Define Groups Define Groups | Group <u>1</u> : 1 Group <u>2</u> : 2 |
| | Continue Cancel Help |
| Moses extreme reactions Wald-Wolfowitz runs OK Reset Cancel Help | |

The Output is:



The conclusion is, that on the basis of this sample, there is evidence to suggest that the population median highest year of schoolc for males and females are not the same.

Compare this with the t-test result. The probabilities are different, but the conclusion is the same.

Wilcoxon Signed-Ranks test for paired samples

We shall again use the SPSS data set New drug.sav for this example. This is a very small data set but we shall assume the subjects were randomly selected.

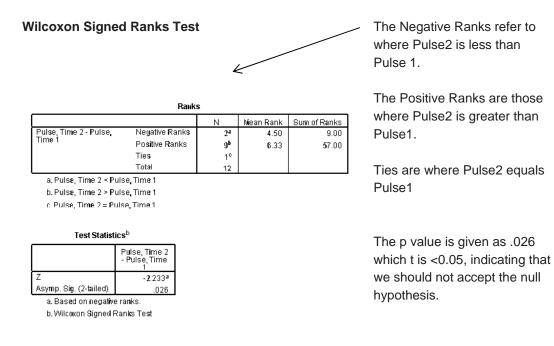
| The question: | Is there a difference in the population median of pulse rates 1 and 2 of patients. |
|--------------------------|---|
| The Research Hypothesis: | There is a difference in the population median of pulse rates 1 and 2 of patients. |
| The Null Hypothesis: | There is no difference in the population median of pulse rates 1 and 2 of patients. |

We are comparing the differences between pairs of readings that are related: the two pulse rates are from the same patient.

Use Analyze > Nonparametric Tests > 2 Related Samples

| | Two-Related-Samples Tests | × |
|--|---|-----------------------------------|
| Complete the dialogue box | Prug [drug] Pair Variable1 Variable2 Arrow Time 2 free 2 | Exact ŷ Options |
| by placing both Pulse, Time1 and Pulse, Time2 | Respiratory, Time 3 [res. Pulso, Time 1 [pulso1] Pulso, Time 1 [pulso2] Pulso, Time 2 [pulse2] Pulso, Time 2 [pulse2] | ↓ |
| in the Test Pairs box | Test Type V Milcoxon Sign | |
| and ticking the <u>W</u> ilcoxon box. | CK Paste Reset Cancel Help | |

The Output is:



The conclusion is that there is a difference in the two pulse rates of the patients.

8. And finally

This is not a statistics textbook. This has been a book about using SPSS, written for non statisticians.

You are probably reading it because you have data to analyse, and want to find out how SPSS can help you. It won't be able to help unless you understand what your data is measuring, which of your numbers mean a measurement, and which are merely shorthand codes for answering "Yes" or "I do a lot of training."

Time spent thinking about your data is never wasted. Think about what you would like your final report to say; it will direct your analysis. Firstly though do the simple stuff: look at frequencies, draw charts (simple ones) and produce two way tables. Make sure you produce two of these each time, one showing row percentages and one showing column percentages, and don't be tempted to combine them in one because that leads to confusion. Keep it simple.

With luck through doing this the data should start to tell you its story, and once you have a handle on that you will be well away.

Because this is not statistics textbook I suggest you find one that suits you and consult it from time to time. Better still find a statistician, who will be very grateful for all the simple stuff you have done first!

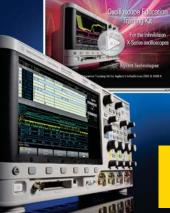
Budget-Friendly. Knowledge-Rich.

The Agilent InfiniiVision X-Series and 1000 Series offer affordable oscilloscopes for your labs. Plus resources such as lab guides, experiments, and more, to help enrich your curriculum and make your job easier.



Scan for free Agilent iPhone Apps or visit qrs.ly/po20pli

© Agilent Technologies, Inc. 2012





u.s. 1-800-829-4444 canada: 1-877-894-4414

Anticipate ____Accelerate ____Achieve



Agilent Technologies